AD 741760

ARL 72-0012
JANUARY 1972

Aerospace Research Laboratories

WIDELY USABLE UNCONDITIONAL MODELS FOR ANALYSIS OF VARIANCE WITH FIXED OR RANDOM EFFECTS

JOHN E. WALSH

SOUTHERN METHODIST UNIVERSITY
DALLAS, TEXAS

CONTRACT NO. F33615-71-C-1178
PROJECT NO. 7071

Approved for public release; distribution unlimited.

AIR FORCE SYSTEMS COMMAND
United States Air Force

NOTICES

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Agencies of the Department of Defense, qualified contractors and other government agencies may obtain copies from the

Defense Documentation Center
Cameron Station
Alexandria, Virginia 22314

This document has been released to the

CLEARINGHOUSE
U. S. Department of Commerce
Springfield, Virginia 22151

for sale to the public.

Copies of ARL Technical Documentary Reports should not be returned to Aerospace Research Laboratories unless return is required by security considerations, contractual obligations or notices on a specified document.

AIR FORCE: 25-4-72/600

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Southern Methodist University Dallas, Texas 75222	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

Widely Usable Unconditional Models for Analysis of Variance With Fixed or Random Effects

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Scientific Interim.

5. AUTHOR(S) (First name, middle initial, last name)

John E. Walsh

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
January 1972	70	6

8a. CONTRACT OR GRANT NO. F33615-71-C-1178 b. PROJECT NO. 7071-00-10 c. DoD Element 61102F d. DoD Subelement 681304	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) ARL 72-0012

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
TECH OTHER	Aerospace Research Laboratories (LB) Air Force Systems Command Wright-Patterson AFB, Ohio 45433

13. ABSTRACT

Many results for statistical investigations are valid for more general models than were used in their development. This is the case for situations where independence (or zero correlation), equal variances, and sometimes joint normality are assumed for sets of random variables. This research investigates extended uses of standard results for some analysis of variance situations. Many methods are found to be usable with much more general models, although not all types of effects can be investigated. The research also provides investigation methods that are usable for all the effects. Moreover, the extended models associated with these methods are much more general than those for the standard methods. Chapter I is introductory and includes a statement of the concepts involved. Material in Chapter II shows that some of the results assuming a sample from a normal population are applicable more generally, including some methods for one-way analysis of variance. The remaining chapters are, with one exception, concerned entirely with analysis of variance. Chapter III contains some extended uses of standard results for fixed-effects models. Chapter IV gives extended uses of standard results for models with random effects. Chapter V contains methods for investigating all effects, for some models with fixed effects. Chapter VI is concerned with investigating all effects and models with random effects. Finally, an analysis of covariance model and a mixed model are considered in Chapter VII.

DD FORM 1 NOV 65 1473

Unclassified
Security Classification

Unclassified
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Linear models						
Analysis of variance						
Variance components						
Robustness						

☆U.S.Government Printing Office: 1972 — 759-084/500

Unclassified
Security Classification

ARL 72-0012

WIDELY USABLE UNCONDITIONAL MODELS FOR ANALYSIS OF VARIANCE WITH FIXED OR RANDOM EFFECTS

JOHN E. WALSH

SOUTHERN METHODIST UNIVERSITY
DALLAS, TEXAS

JANUARY 1972

CONTRACT NO. AF33615-71-C-1178
PROJECT NO. 7071

AEROSPACE RESEARCH LABORATORIES
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

FOREWORD

This research was sponsored by the Aerospace Research Laboratories, Air Force Systems Command, United States Air Force, Wright-Patterson Air Force Base, Ohio. The work was performed under Contract F33615-71-C-1178, which was technically monitored by Dr. David A. Harville, and was a part of Project 7071, Research in Applied Mathematics.

ABSTRACT

Many results for statistical investigations are valid for more general models than were used in their development. This is the case for situations where independence (or zero correlation), equal variances, and sometimes joint normality are assumed for sets of random variables. This research investigates extended uses of standard results for some analysis of variance situations. Many methods are found to be usable with much more general models, although not all types of effects can be investigated. The research also provides investigation methods that are usable for all the effects. Moreover, the extended models associated with these methods are much more general than those for the standard methods. Chapter I is introductory and includes a statement of the concepts involved. Material in Chapter II shows that some of the results assuming a sample from a normal population are applicable more generally, including some methods for one-way analysis of variance. The remaining chapters are, with one exception, concerned entirely with analysis of variance. Chapter III contains some extended uses of standard results for fixed-effects models. Chapter IV gives extended uses of standard results for models with random effects. Chapter V contains methods for investigating all effects, for some models with fixed effects. Chapter VI is concerned with investigating all effects and models with random effects. Finally, an analysis of covariance model and a mixed model are considered in Chapter VII.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

Models for observed data, in terms of parameters and unobserved random variables, provide the basis for investigations of an analysis of variance (ANOVA) nature. The customary models utilize some stringent assumptions, which include equal variances within sets for the unobserved random variables (perhaps for all such random variables) and zero correlation among most or all of the unobserved random variables. Sometimes, joint normality for the unobserved random variables is also assumed (when something besides point estimation is desired).

Many methods for statistical investigation are applicable for more general models than those on which their development was originally based. The research reported here identifies extended models for which some standard ANOVA type methods remain applicable. These extended models are substantial generalizations of the original models. However, no extension occurs wherein all of the types of effects can be investigated by use of standard methods.

To obtain investigation methods for all the types of effects, and to be able to use extended models of a much more general nature, some statistical methods different from those customarily used are developed. This is accomplished by replacing the usual "sum of squares for error" by another statistic that has similar properties but fewer "degrees of freedom."

An extended model is obtained by adding one or more unobserved random terms of an error nature to the original model. The kind of terms added, and conditions they are required to satisfy, depend on the types of effects that are to be investigated simultaneously. An effect investigated is a parameter in the model or, for random effects, is a variance (or covariance, or function of variances and/or covariances) of the type of unobserved random effects, in the original model,

being considered. There is no interest in properties of the extra random error(s) that are added to the original model to form the extension.

The generality level for an extended model depends on the types of effects to be investigated simultaneously. For the extensions given, the generality level is much higher for a subset of a set (of types of effects) than it is for the set itself.

A requirement in the development of an extended model is that the additional error term(s) must not occur in any of the statistics used for the investigation methods. That is, the extensions are motivated by the statistics to be used. Two ways, in combination, are applied to eliminate the extra error terms from statistics. First, these errors can be exactly cancelled out in the statistics. Second, one or more summation conditions are imposed that cause these errors to be eliminated. The summation conditions reduce the level of generality obtained from the presence of the additional errors but much of this generality remains. For example, the requirement that ten random errors sum to a constant effectively leaves nine unconstrained random error terms. Since all additional random errors are eliminated, the statistics have the same probability properties for the extension as they do for the original model. In particular this is the case when the joint normality assumption occurs for the original model.

To be emphasized is that the total interest is concentrated on investigation of the effects that occur for the original model. Any additional random error term is a further error contribution from the experimental situation (and ideally would be zero).

Also to be emphasized is that, subject to any summation conditions imposed, the additional random error terms can have an arbitrary joint distribution. That is, they can have any joint distribution that is possible. Moreover, the additional errors can have any permissible dependences with the unobserved random terms in the original model.

When only one extra error term occurs, it can have an arbitrary distribution and any permissible dependences with unobserved random variables in the original model.

Thus, an extended model is able to at least approximate a broad class of joint distributions for the observations. Results obtained for an extended model are exact when some one of the possibilities for the extended model exactly represents the joint distribution of the observed random variables. Approximate results occur when an approximate (but not exact) representation is obtainable.

These considerations, and the research performed, show that many of the standard results for models of an ANOVA nature have "robustness" properties. That is, for the models considered, many of the effects investigated for the original model are still investigated, with the same probability properties as for the standard methods, when additional error terms of a rather general nature can be present. Moreover, modified investigation methods combined with further research show that (for the models considered) all effects can be investigated and with stronger robustness properties than for the standard methods.

The first material presented (Chapter II) considers extensions for the cases where a random sample from a normal population is originally assumed, and for the case where the several-sample normality model for one-way ANOVA is originally assumed. Standard methods are considered for both cases. This first material is given in greater detail than for other cases using standard methods, in order to introduce the cancellation way of eliminating the extra error terms.

The second material (Chapter III) considers extensions of models with fixed effects for use of the standard methods. Two-way ANOVA with no replication and two-way ANOVA with replication (interactions included) are the two cases examined. Some results of this nature are also given in Chapter V (for presentation convenience).

The third set of material (Chapter IV) is concerned with the extensions of

random-effects (variance components) models and use of the customary methods. Considered in somewhat general terms are various types of variance components models. Considered in detail is two-way ANOVA with no replication.

Chapter V, containing the fourth set of material, considers extensions of fixed-effects models and some new investigation methods. Here, all effects can be investigated. One-way ANOVA and two-way ANOVA with replication (and interactions) receive consideration. More detail than usual (for cases with new methods) occurs here, in order to introduce the summation way of eliminating extra error terms.

The fifth material, in Chapter VI, is concerned with extensions of random-effects models and presents some new methods that can be used to investigate all effects. One-way ANOVA and twofold nested ANOVA are considered.

Finally, in Chapter VII, some results for one-way analysis of covariance and for the Scheffé two-way mixed model for ANOVA are presented. Standard methods are considered for the analysis of covariance case, and new methods are given for the mixed model case. All effects can be investigated for the mixed model case.

In addition to the basic investigation results, several methods have been developed for rejection of outlying observations. For brevity, these methods are not presented in all cases. However, to illustrate the approach used, a few of the methods are given in detail. Also, other places where such a method has been developed are identified by specification of the most general extension (of the type considered) for which the rejection-method is applicable.

Lastly, some notation properties are stated. This notation involves the use of a dot in place of a subscript for an observed variable (random, or fixed as in the analysis of covariance). Use of a dot in place of a subscript implies that the quantity considered is the arithmetic average of the variable over all values of this subscript, for the specified values of the other subscripts. For example, consider y_{ijk}, where $i = 1, \ldots, I$; $j = 1, \ldots, J$; $k=1, \ldots, K$.

$$y_{ij.} = \sum_{k=1}^{K} y_{ijk}/K, \qquad y_{..k} = \sum_{i=1}^{I} \sum_{j=1}^{J} y_{ijk}/IJ,$$

$$y_{...} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} y_{ijk}/IJK.$$

This notation property is used throughout Chapters II - VII. Also, the way of representing equation numbers and numbers for ANOVA type models is selfcontained within each part that begins 0. SUMMARY, or begins 0 . Summary. That is, the numbering used applies only within this part.

CHAPTER II

EXACT NORMALITY-SAMPLE RESULTS FOR DEPENDENT OBSERVATIONS FROM NONNORMAL POPULATIONS

0. SUMMARY

Some well-known statistical results are based on the assumption of a random sample from a normal population or, for one-way ANOVA, the assumption of independent samples from normal populations. Many of these tests, confidence regions, etc. have the same properties when less stringent assumptions are made (and the investigation is suitably interpreted). That is, a generalization can be made of the concept of a sample from a normal population and of the concept of independent samples from normal populations. These generalized models permit dependence among some or all observations and each observation can be from a different nonnormal population. Exact results are obtainable when the statistic(s) used can be expressed entirely in terms of differences of observations. For the situation corresponding to the one-sample case, such statistics occur for variance investigation and rejection of outliers. Some statistics of this nature occur for the situation corresponding to one-way ANOVA, including the F-statistic often used to test whether independent samples are from the same normal population.

1. INTRODUCTION AND RESULTS

A common assumption is that the data are a random sample from a normal population or, for one-way ANOVA, are independent samples from normal populations. More specifically, for the one-sample case, the observations can be expressed in the form

$$x_i = \mu + e_i, \qquad (i = 1,\ldots,n),$$

where the e_i are a random sample from a normal population with zero mean and positive variance σ^2 (ordinarily unknown). For the one-way ANOVA case, the observations can be expressed in the form

$$x_{ij} = \mu_j + e_{ij}, \qquad [i = 1,\ldots,n(j);\ j = 1,\ldots,m],$$

where, for fixed j, the e_{ij} are a random sample of size n(j) from a normal population with zero mean and positive variance σ_j^2 (ordinarily unknown). Also, the e_{ij} are mutually independent.

The purpose of this material is to show that many of the results developed on the basis of these normality-sample assumptions remain exact, and have the same properties, when the forms used for the observations are generalized.

Corresponding to the one-sample case, the observations are expressed in the form

$$x_i = \mu + e_i + e', \qquad (i = 1,\ldots,n),$$

where the e_i have the same properties as for the normality-sample situation and e' has an arbitrary distribution. Also, e' can have any allowable dependence with the e_i and the level of dependence can vary with i. If the variances and covariance exist for $(e', e_1, \ldots, e_n)$, with positive variance for e', and ρ_i denotes the correlation between e' and e_i, the condition

$$\rho_1^2 + \ldots + \rho_n^2 \le 1 \tag{1}$$

must be satisfied by the ρ_i.

The interest is still in the properties of the population yielding the e_i, with no interest in properties that involve e'. The value of e' represents an error contribution imposed on all the observations by the experimental situation (and ideally would be zero), with different e_i possibly having a different influence on the random value occurring for e'. The x_i have continuous distributions, since the e_i have continuous distributions, and have the same population mean (when the expected value of e' exists). Since the distribution of $e_i + e'$ is influenced by the dependence between e_i and e', the x_i can have noticeably different nonnormal distributions. Also, any two x_i are virtually always dependent, and the

level of dependence can be high. A special case is that where the x_i are normal multivariate and the correlation between every two x_i is the same. Properties of some well-known results for this case are examined in Walsh (1947). As is shown in the verification section, the model for the x_i used here is equivalent to that of Walsh (1947) for the case where $(e', e_1, \ldots, e_n)$ have a normal multivariate distribution and the β_i are equal. The generalized form used for the x_i, also the form given later for the x_{ij}, are specializations of the form introduced in Walsh (1968) for two-way ANOVA.

When a statistic can be expressed entirely in terms of differences of the x_i, the contribution e' cancels out and the statistic has the same properties as when the normality-sample assumptions hold. As an example, for $n \geq 2$,

$$s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2, \text{ with } \bar{x} = \sum_{i=1}^{n} x_i/n$$

identically equals

$$\sum_{i=1}^{n} (e_i - \bar{e})^2, \text{ with } \bar{e} = \sum_{i=1}^{n} e_i/n.$$

Hence, s^2/σ^2 has a χ^2 - distribution with $n - 1$ degrees of freedom and $s^2/(n-1)$ is an exactly unbiased estimate of σ^2. Also, procedures for rejection of outlying observations almost always are based on differences of the sample values (for example, see Guttman and Smith (1969), and the references listed therein).

Next, consider the form of the observations for the situation that corresponds to one-way ANOVA for two or more samples. Here,

$$x_{ij} = \mu_j + e_{ij} + e'', \quad [i = 1, \ldots, n(j);\ j = 1 \ldots, m],$$

where the e_{ij} have the same properties as for the normality-sample case and the distribution of e'' is arbitrary. Also, e'' can have any permissible dependence with the e_{ij}, and the level can vary with the value of (i,j). When variances and

covariances exist for $(e'', e_{ij}; i=1,\ldots,n(j);\ j=1,\ldots,m)$, ρ_{ij} denotes the correlation coefficient for e'' and e_{ij}, the condition

$$\sum_{j=1}^{m} \sum_{i=1}^{n(j)} \rho_{ij}^{2} \le 1 \qquad (2)$$

is necessarily satisfied by the ρ_{ij}.

The only interest is in the properties of the populations from which the e_{ij} are obtained. The value of e'' is an error contribution that ideally would be zero. The null hypothesis considered is the same as for the normality sample case. That is, the null hypothesis asserts that the $e_{ij} + \mu_j$ are from the same normal population. In view of this null hypothesis, and reasons for its investigation, consideration of the case where all observations receive the contribution e'' seems to be an appropriate generalization of the form of the observations for one-way ANOVA. The distributions of the x_{ij} are continuous but can be nonnormal and noticeably different. Any two x_{ij} are virtually always dependent, and this dependence can be very strong.

The contribution e'' cancels out for any statistic that can be expressed entirely in terms of differences of the x_{ij}. In particular, suppose that all $n(j) \ge 2$ and consider the F-statistic often used to test whether independent samples are from the same normal population. Under the null hypothesis, the $\mu_j + e_{ij}$ are from the same normal population, so that

$$S_E^{2} = \sum_{j=1}^{m} \sum_{i=1}^{n(j)} (s_{ij}-\bar{x})^{2} = \sum_{j=1}^{m} \sum_{i=1}^{n(j)} (e_{ij}-\bar{e})^{2},$$

$$S_B^{2} = \sum_{j=1}^{m} n(j)\,(\bar{x}_{.j}-\bar{x})^{2} = \sum_{j=1}^{m} n(j)\,(\bar{e}_{.j}-\bar{e})^{2},$$

where

$$\bar{x} = \sum_{j=1}^{m} \sum_{i=1}^{n(j)} x_{ij}/N, \qquad \bar{e} = \sum_{j=1}^{m} \sum_{i=1}^{n(j)} e_{ij}/N,$$

$$\bar{x}_{\cdot j} = \sum_{i=1}^{n(j)} x_{ij}/n(j), \qquad \bar{e}_{\cdot j} = \sum_{i=1}^{n(j)} e_{ij}/n(j),$$

and

$$N = \sum_{j=1}^{m} n(j).$$

Thus, the statistic

$$(N-m)\, S_B^{\;2}/(m-1)\, S_E^{\;2}$$

has an F-distribution with m-1 and N-m degrees of freedom when the null hypothesis holds. In general, the distribution of this statistic is the same as for the normality-sample case of one-way ANOVA.

2. VERIFICATIONS

First, consider verification of Equation (1) for the situation corresponding to the one-sample case where the variances and covariances exist for $(e', e_1, \ldots, e_n)$ and all variances are positive. The determinant of the variance-covariance matrix for $(e', e_1, \ldots, e_n)$ can be expressed as

$$\begin{vmatrix} \sigma_0^{\;2} & \rho_1\sigma_0\sigma & \rho_2\sigma_0\sigma & \cdots & \rho_n\sigma_0\sigma \\ \rho_1\sigma_0\sigma & \sigma^2 & 0 & \cdots & 0 \\ \rho_2\sigma_0\sigma & 0 & \sigma^2 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_n\sigma_0\sigma & 0 & 0 & \cdots & \sigma^2 \end{vmatrix}$$

with $\sigma_0^{\ 2}$ being the variance of e'. Expansion shows that its value is

$$\sigma_0^{\ 2}\sigma^{2n}(1-\rho_1^{\ 2}-\ldots-\rho_n^{\ 2}),$$

which is necessarily nonnegative.

Next, consider verification of Equation (2) for the situation corresponding to the one-way ANOVA case where the variances and covariances exist for $(e'', e_{ij};\ i=1,\ldots,n(j);\ j=1,\ldots,m)$ and all variances are positive. The determinant of the variance-covariance matrix for this multivariate random variable is

$$\begin{vmatrix} \sigma_{00}^{\ 2} & \rho_{11}\sigma_{00}\sigma_1 & \cdots & \rho_{n(m)m}\rho_{00}\rho_m \\ \rho_{11}\sigma_{00}\sigma_1 & \sigma_1^{\ 2} & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \rho_{n(m)m}\sigma_{00}\sigma_m & 0 & \cdots & \sigma_m^{\ 2} \end{vmatrix}$$

with $\sigma_{00}^{\ 2}$ being the variance of e''. The value of this determinant is

$$\sigma_{00}^{\ 2}\sigma_1^{\ 2n(1)}\ldots\sigma_m^{\ 2n(m)}\left[1-\sum_{j=1}^{m}\sum_{i=1}^{n(j)}\rho_{ij}^{\ 2}\right],$$

which is necessarily nonnegative.

Finally, suppose that $(e', e_1, \ldots, e_n)$ has a normal (n+1)-variate distribution with $\sigma_0^{\ 2}$ and σ^2 positive and

$$\rho_1 = \ldots = \rho_n = \rho.$$

Then the correlation between any two x_i is the same and can be expressed as

$$(\sigma_0^{\ 2}+\rho\sigma_0\sigma)/(\sigma_0^{\ 2}+\sigma^2+2\rho\sigma_0\sigma). \qquad (3)$$

The model here for the one-sample case is equivalent to the model in Walsh (1947) if the value of (3) can be shown to take all values from $-(n-1)^{-1}$ to 1. Since (3) is

continuous in σ_0/σ and ρ, it is sufficient to show that $-1/(n-1)$ and 1 are attainable (at least in the limit). First, the value of (3) becomes 1 as $\sigma_0/\sigma \to \infty$. Second, from Equation (1),

$$-n^{-\frac{1}{2}} \leq \rho \leq n^{\frac{1}{2}}.$$

Let $\rho = -n^{-\frac{1}{2}}$ and $\sigma_0/\sigma = n^{\frac{1}{2}}$. Then, the value of (3) is $-(n-1)^{-1}$. Thus, these two models are equivalent.

CHAPTER III

EXTENDED USES OF STANDARD RESULTS WITH FIXED EFFECTS

TWO-WAY ANOVA WITH NO REPLICATION

0. Summary

Consider the standard two-way cross-classification model for ANOVA with one observation for each combination of factor levels and fixed effects. The model with no interaction is used. The customary assumption of joint normality is made for random terms when procedures other than point estimation are used. This standard model is generalized, in several ways, by addition of one or two more random terms of an "error" nature. The extended models are much more generally usable than the standard models. However, except when all types of effects are to be investigated simultaneously, a customary procedure for the standard model remains applicable for one or more of its extensions.

1. Introduction

The standard model for two-way ANOVA with cross-classification, no interaction, fixed effects, and no replication is

$$y_{jk} = \mu + \alpha_j + \beta_k + e_{jk}, \tag{1}$$

where $j = 1,\ldots,J$ and $k = 1,\ldots, K$, with $J,K \geq 2$. Here, y_{jk} is an observed random variable, μ is a parameter, α_j is a parameter satisfying $\alpha_1 + \ldots + \alpha_J = 0$, and β_k is a parameter satisfying $\beta_1 + \ldots + \beta_K = 0$. The e_{jk} are unobserved random variables that have zero mean, the same variance σ^2, and are mutually uncorrelated. The e_{jk} are also assumed to have a joint normal distribution when significance tests or confidence regions are desired. Some or all of μ, σ^2, one or more of the α_j, and one or more of the β_k can be investigated when model (1) applies.

This material shows that, except when all types of effects are investigated,

extensions occur where the standard investigation methods remain applicable. Seven extensions are given. As always, the total interest is in investigating the effects for model (1), with no interest in examining properties of the extra random error terms added to this model.

2. The Extensions

An extension is identified by the one or two random terms that are added to the expression for y_{jk} in the standard model, and by the allowable probability properties for the additional random term(s).

First, consider the case where σ^2, α_j, β_k are the types of effects investigated. The extension consists of

$$y_{jk} = (\text{standard model}) + e', \qquad (2)$$

where, as for all the extensions, the quantities occurring in the standard model have the same properties as for that model. The additional random error e' can have an arbitrary distribution, and can have any allowable dependences with the random error terms in model (1).

Second, the types of effects investigated are μ, σ^2, α_j and the extended model is

$$y_{jk} = (\text{standard model}) + e'_k, \qquad (3)$$

where the additional random errors satisfy $e'_1 + \ldots + e'_K = 0$. Otherwise, the e'_k can have an arbitrary joint distribution, and any permissible dependences with the random errors in the standard model.

Third, μ, σ^2, β_k are the types investigated and the extension is

$$y_{jk} = (\text{standard model}) + e''_j, \qquad (4)$$

where the additional random errors satisfy $e''_1 + \ldots + e''_J = 0$.

Fourth, the types investigated are μ, σ^2 and the extended model is

$$y_{jk} = (\text{standard model}) + e'_k + e''_j, \qquad (5)$$

where the additional random errors satisfy

$$\sum_{j=1}^{J} \sum_{k=1}^{K} (e'_k + e''_j) = 0.$$

Model (5) is an extension of models (3) and (4).

Fifth, σ^2 and α_j are the types investigated and the extension is

$$y_{jk} = \text{(standard model)} + e'_k. \tag{6}$$

Model (6) is an extension of models (2) and (3).

Sixth, σ^2 and β_k are the types investigated and the extended model is

$$y_{jk} = \text{(standard model)} + e''_j. \tag{7}$$

Model (7) is an extension of models (2) and (4).

Finally, consider the case where σ^2 is the type of effect investigated. The extension is

$$y_{jk} = \text{(standard model)} + e'_k + e''_j. \tag{8}$$

Model (8) is an extension of all the other extended models.

3. Verification Outline

First, the standard statistics for model (1) are stated. Then, for each appropriate type of statistic, or combination of types, the extension (or extensions) for which the extra random errors do not occur are identified. For a given investigation, the only usable extensions are those where the additional random errors are eliminated in all the statistics for the investigation. Then, the probability properties for the statistics are the same as for model (1), so that the customary results apply (for example, see Graybill (1961)).

The statistics considered are

$$\hat{\mu} = y_{..}, \quad \hat{\alpha}_j = y_{j.} - y_{..}, \quad \hat{\beta}_k = y_{.k} - y_{..},$$

$$s_a^2 = \sum_{j=1}^{J} (y_{j.} - y_{..})^2, \qquad s_b^2 = \sum_{k=1}^{K} (y_{.k} - y_{..})^2,$$

$$s_e^2 = \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{jk} - y_{j.} - y_{.k} + y_{..})^2.$$

Here, s_e^2 is used to investigate σ^2. Also, s_e^2 is used in investigations for the other types of effects (tests, confidence regions, estimates of variances of estimates). The statistic s_e^2 is free of the additional random terms for all of the extensions.

The statistics $\hat{\alpha}_j$ and s_a^2 are used to investigate the a_j. Both s_a^2 and the $\hat{\alpha}_j$ are free of the additional random terms for models (2), (3), and (6).

The statistics $\hat{\beta}_k$ and s_b^2 are used to investigate the β_k. Both s_b^2 and the $\hat{\beta}_k$ are free of the additional random variables for models (2), (4), and (7).

Finally, the statistic $\hat{\mu}$ is used to investigate μ, and is an unbiased estimate of μ. The statistic $\hat{\mu}$ is free of the additional random variables for models (3), (4), and (5).

TWO-WAY ANOVA WITH INTERACTIONS

0. Summary

Considered is the standard two-way cross-classification model, with fixed effects and including interactions. This standard model is extended by addition of one or two more "error" terms. However, except when all the types of effects are to be considered simultaneously, the customary procedures for the standard model remain applicable for at least one of the extended models. Also, a procedure for rejection of outliers using the standard model is outlined, and shown to be applicable for the most general extension.

1. Introduction

The balanced model for two-way ANOVA, with cross-classification, interactions, and fixed effects is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}, \tag{1}$$

where $i = 1,\ldots,I$; $j = 1,\ldots,J$; $k = 1,\ldots,K$; with $I, J, K \geq 2$. Here, y_{ijk} is an observed random variable, μ is a parameter, α_i is a parameter satisfying $\alpha_1 + \ldots + \alpha_I = 0$, and β_j is a parameter satisfying $\beta_1 + \ldots \beta_J = 0$. The γ_{ij} are parameters that satisfy $\gamma_{i1} + \ldots + \gamma_{iJ} = 0$ for all i, and $\gamma_{1j} + \ldots + \gamma_{Ij} = 0$ for all j. The e_{ijk} are unobserved random variables and are mutually uncorrelated with zero mean and variance σ^2. The e_{ijk} are also assumed to have a joint normal distribution when something besides point estimation is desired.

First, consider a description of a method for deciding whether an observation is an outlier. Let y_{ijk*} be selected, without knowledge of the observation values, for investigation as a possible outlier. Divide the observations with this i, j value into sets of size two, and one or zero sets of size three (unbiasedly) so that y_{ijk*} is not in a set of size three. When a set of size three occurs, it is modified to a set of size two by (unbiasedly) combining two of its observations. Specifically, two of the observations are added and their sum divided by $\sqrt{2}$, to yield one "observation." This "observation" is identified by using the larger of the two values for k in the observations summed. For each of the resulting sets, a statistic of the kind

$$y_{ijk(1)} - y_{ijk(2)} = e_{ijk(1)} - e_{ijk(2)}$$

is formed, where $k(1) = k*$ for the set containing y_{ijk*}. These statistics are uncorrelated with the same variance and zero mean. The statistic containing y_{ijk*} can be investigated as an outlier, when the assumption of normality for the e_{ijk} is also made. This can be done by use of a procedure for deciding whether a specified observation, believed to be in a sample from a normal population with zero mean (and selected without knowledge of the sample values), is an outlier.

Fifteen extensions of model (1) are given. An extension is made by addition of one, or two, more "error" terms (of different kinds) to the standard model for y_{ijk}.

2. Extensions

No extension occurs for the case where μ, σ^2, at least one of the α_i, the least one of the β_j, and at least one of the γ_{ij} are all to be investigated. However, extended models are given for the other important cases.

First, consider the extension where the types of effects investigated are σ^2, α_i, β_j, γ_{ij}. The extended model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e', \qquad (2)$$

where, as for all the extensions, μ, α_i, β_j, γ_{ij}, and e_{ijk} have the same properties as in model (1).

Second, the types of effects investigated are μ, σ^2, α_i, γ_{ij} and the extension is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e'_j, \qquad (3)$$

where the additional random errors satisfy $e'_1 + \ldots + e'_J = 0$.

Third, consider the case where the types of effects investigated are μ, σ^2, β_j, γ_{ij}. The extension as

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e''_i, \qquad (4)$$

where the additional random errors satisfy $e''_1 + \ldots + e''_I = 0$.

Fourth, the types of effects investigated are σ^2, α_i, γ_{ij} and the extended model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ijk} + e_{ijk} + e'_j. \qquad (5)$$

Model (5) is an extension of models (2) and (3).

Fifth, σ^2, β_j, γ_{ij} are the types of effects investigated and the extension is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e''_i. \qquad (6)$$

Model (6) is an extension of models (2) and (4).

Sixth, the types investigated are μ, σ^2, γ_{ij} and the extension is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e'_j + e''_i, \qquad (7)$$

where the additional random variables satisfy

$$\sum_{i=1}^{I} \sum_{j=1}^{J} (e'_j + e''_i) = 0.$$

Model (7) is an extension of models (3) and (4).

Seventh, σ^2, γ_{ij} are the types investigated and the extended model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e'_j + e''_i \ . \qquad (8)$$

Model (8) is an extension of models (5), (6), and (7).

Eighth, the types of effects investigated are μ, σ^2, α_i, β_j and the extension is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e^*_{ij}, \qquad (9)$$

where the additional random variables satisfy ($i = 1,\ldots,I$; $j=1,\ldots,J$)

$$\sum_{i=1}^{I} e^*_{ij} = 0, \qquad \sum_{j=1}^{J} e^*_{ij} = 0.$$

Ninth, μ, σ^2, α_i are the types investigated and the extended model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e^*_{ij}, \qquad (10)$$

where the additional random variables satisfy

$$\sum_{j=1}^{J} e^*_{ij} = 0, \qquad (i=1,\ldots,I).$$

Model (10) is an extension of model (9).

Tenth, the types investigated are μ, σ^2, β_j and the extension is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e^*_{ij}, \qquad (11)$$

where the additional random variables satisfy

$$\sum_{i=1}^{I} e^*_{ij} = 0, \qquad (j=1,\ldots,J).$$

Model (11) is an extension of model (9).

Eleventh, α_i, β_j, σ^2 are the types investigated and the extended model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e^*_{ij}, \qquad (12)$$

where the additional random variables satisfy $(i = 1,\ldots,I;\ j=1,\ldots,J)$

$$\sum_{i=1}^{I} e^*_{ij} = C_1, \qquad \sum_{j=1}^{J} e^*_{ij} = C_2,$$

with C_1, C_2 constants that can have any values. Model (12) is an extension of models (2) and (9).

Twelfth, the types investigated are σ^2, α_i and the extension is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e^*_{ij}, \qquad (13)$$

where the additional random variables satisfy

$$\sum_{j=1}^{J} e^*_{ij} = C, \qquad (i = 1,\ldots,I),$$

with C a constant that can have any value. Model (13) is an extension of models (5), (10), and (12).

Thirteenth, σ^2, β_j are the types investigated and the extension is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e^*_{ij}, \qquad (14)$$

where the additional random variables satisfy

$$\sum_{i=1}^{I} e^*_{ij} = C, \qquad (j=1,\ldots,J).$$

Model (14) is an extension of models (6), (11), and (12).

Fourteenth, the types investigated are μ, σ^2 and the extended model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e^*_{ij}, \qquad (15)$$

where the additional random variables satisfy

$$\sum_{i=1}^{I} \sum_{j=1}^{J} e^*_{ij} = 0.$$

Model (15) is an extension of models (7), (10), and (11).

Finally, σ^2 is investigated and the extended model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e^*_{ij}. \qquad (16)$$

Model (16) is an extension of all the other extensions and, as is easily verified, the method outlined for rejection of outliers (using the standard model) remains applicable for model (16).

3. Outline of Verifications

The statistics customarily used for model (1) are stated and the extended models for which the additional random variables do not occur are identified. Also, identification is given for the type(s) of effects investigated by a stated statistic.

Some further notation is introduced for stating the statistics considered. This is

$$\hat{\mu} = y_{...}, \quad \hat{\alpha}_i = y_{i..} - y_{...}, \quad \hat{\beta}_j = y_{.j.} - y_{...},$$

$$\hat{\gamma}_{ij} = y_{ij.} - y_{i..} - y_{.j.} + y_{...}, \quad S_A^2 = \sum_{i=1}^{I} (y_{i..} - y_{...})^2$$

$$S_B^2 = \sum_{j=1}^{I} (y_{.j.} - y_{...})^2, \quad S_G^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij.} - y_{i..} - y_{.j.} + y_{...})^2,$$

$$S_E^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{ijk} - y_{ij.})^2.$$

The statistic S_E^2 is used to investigate σ^2 and is free of the additional random variables for all the extended models. Also, S_E^2 is used for investigations of the other types of effects (and occurs in significance tests, confidence regions, estimates of variances of estimates, etc.).

The statistics $\hat{\alpha}_i$ and S_A^2 are used to investigate the α_i and are free of the additional random terms for models (2), (3), (5), (9), (10), (12), and (13). The statistics $\hat{\beta}_j$ and S_B^2 are used to investigate the β_j and additional random variables do not occur in these statistics for models (2), (4), (6), (9), (11), (12), and (14).

The statistics $\hat{\gamma}_{ij}$ and S_G^2 are used for investigating the γ_{ij} and are free of additional random variables for models (2) - (8). The statistic $\hat{\mu}$ is used to investigate μ and is free of the additional random terms for models (3), (4), (7), (9), (10), (11), and (15).

CHAPTER IV

EXTENDED USES OF STANDARD RESULTS WITH RANDOM EFFECTS

TWO-WAY ANOVA WITH NO REPLICATION

0. Summary

Considered is the standard two-way classification model for ANOVA with random effects and no replication. No interactions occur. This standard model is generalized by addition of one or two more random terms of an "error" nature. However, except when all types of effects are to be investigated simultaneously, a customary procedure for the standard model remains applicable for one or more extensions.

1. Introduction

The standard model for two-way analysis of variance with cross-classification, random effects (variance components model), and no replication is

$$y_{jk} = \mu + a_j + b_k + e_{jk}, \tag{1}$$

where $j = 1,\ldots,J$ and $k = 1,\ldots,K$, with $J,K \geq 2$. Here, y_{jk} is an observed random variable, μ is a parameter, the a_j are unobserved random variables with zero mean and variance σ_a^2, the b_k are unobserved random variables with zero mean and variance σ_b^2, the e_{jk} are unobserved random variables with zero mean and variance σ_e^2, and the a_j, b_k, c_{jk} are mutually uncorrelated. When tests or confidence regions are desired, the a_j, b_k, c_{jk} are also assumed to have a joint normal distribution.

The types of effects receiving consideration are μ, σ_a^2, σ_b^2, and σ_c^2. Except when all these effects are to be investigated simultaneously, extensions are given for which the standard investigation methods are still usable. Seven extended models are stated.

2. The Extended Models

An extended model is identified by the one or two random terms that are added to the expression for y_{jk} in the standard model, and by the allowable probability

properties for the additional random error(s).

First, consider the case where σ_c^2, σ_a^2, σ_b^2 are the types of effects investigated. The extended model consists of

$$y_{jk} = \text{(standard model)} + e', \tag{2}$$

where, as for all the extensions, the quantities that occurred in the standard model have the same properties as for that model.

Second, the types of effects investigated are μ, σ_c^2, σ_a^2 and the extended model is

$$y_{jk} = \text{(standard model)} + e'_k, \tag{3}$$

where the additional random variables satisfy $e'_1 + \dots + e'_K = 0$.

Third, μ, σ_c^2, σ_b^2 are the types investigated and the extension is

$$y_{jk} = \text{(standard model)} + e''_j, \tag{4}$$

where the additional random variables satisfy $e''_1 + \dots + e''_J = 0$.

Fourth, the types investigated are μ and σ_c^2 and the extended model is

$$y_{jk} = \text{(standard model)} + e'_k + e''_j, \tag{5}$$

where the additional random variables satisfy

$$\sum_{j=1}^{J} \sum_{k=1}^{K} (e'_k + e''_j) = 0.$$

Model (5) is an extension of models (3) and (4).

Fifth, σ_c^2 and σ_a^2 are investigated and the extension is

$$y_{jk} = \text{(standard model)} + e'_k. \tag{6}$$

Model (6) is an extension of models (2) and (3).

Sixth, σ_c^2 and σ_b^2 are investigated and the extended model is

$$y_{jk} = \text{(standard model)} + e''_j. \tag{7}$$

Model (7) is an extension of models (2) and (4).

Seventh, consider the case where σ_c^2 is the effect investigated. The extension is

$$y_{jk} = (\text{standard model}) + e'_k + e''_j . \tag{8}$$

Model (8) is an extension of all the other extensions.

3. Outline of Justification

The standard statistics used with model (1) are defined first. Then, these statistics are associated with the extended models to which they are applicable. The statistics are

$$\hat{\mu} = y_{..} ,$$

$$s_a^2 = \sum_{j=1}^{J} (y_{j.} - y_{..})^2 , \qquad s_b^2 = \sum_{k=1}^{K} (y_{.k} - y_{..})^2 ,$$

$$s_c^2 = \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{jk} - y_{j.} - y_{.k} + y_{..})^2 .$$

Here, s_c^2 is used for investigating σ_c^2, and is used in the investigations for other types of effects. The statistic s_c^2 is free of the additional random terms for all the extensions.

The statistic s_a^2 is used to investigate σ_a^2, and is free of the additional random error terms for extended models (2), (3), and (6).

The statistic s_b^2 is used to investigate σ_b^2, and is free of the extra random error terms for extended models (2), (4), and (7).

Finally, $\hat{\mu}$ is used to investigate μ, and is an unbiased estimate of μ. The statistic $\hat{\mu}$ is free of the additional random variables for extended models (3), (4), and (5).

VARIANCE COMPONENTS MODELS-GENERAL

0. Summary

Considered are standard N-way ($N \geq 1$) variance components (random effects) models for ANOVA. These are examined in a somewhat general sense, with interest

in verifying that many of the customary investigation methods also apply for extensions of these models. An extension is obtained by adding a further random error term to the standard model, where the error term sometimes is different for various cases of the model. Two specific illustrations are given. Also, some ways of investigating outliers for nested designs, which also apply to the extensions, are discussed.

1. Introduction

The statistical results that occur for the usual variance components models can also be obtained for generalizations of these models that use substantially less restrictive assumptions.

To illustrate the generalization approach, consider the usual model for a balanced one-way classification. An observation is denoted by y_{ij}, where $i = 1, \ldots, I$ and $j = 1, \ldots, J$. The model assumes that

$$y_{ij} = \mu + a_i + b_{ij}, \tag{1}$$

where μ is a parameter, each a_i is a random variable with zero mean and variance σ_a^2, and each b_{ij} is a random variable with zero mean and variance σ_b^2.

The statistics used for investigating σ_a^2 are

$$\sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij} - y_{i.})^2 \text{ and } \sum_{i=1}^{I} (y_{i.} - y_{..})^2 \tag{2}$$

(for example, see Graybill (1961)).

The extension of model (1) permits the presence of an additional random variable e' and assumes that

$$y_{ij} = \mu + a_i + b_{ij} + e''. \tag{3}$$

The a_i and the b_{ij} have the same properties as for model (1). Model (3) allows both σ_a^2 and σ_b^2 to be investigated. When only σ_b^2 is considered, a much more general model can be used in which e_i is the additional random error in the model

for y_{ij} $(i = 1, \ldots, I)$.

It is to be noted that, for the normality case of this nested design, models (1), (3) and that for investigating σ_b^2 allow investigation of whether the value of a specified b_{ij} is an outlier. With I even, given i, and given $j = j(1)$, an investigation is based on

$$y_{ij(1)} - y_{ij(2)}, y_{ij(3)} - y_{ij(4)}, \ldots, y_{ij(I-1)} - y_{ij(I)},$$

where $j(1),\ldots,j(I)$ are the different values for j. These differences are independent and equal

$$b_{ij(1)} - b_{ij(2)}, \ldots, b_{ij(I-1)} - b_{ij(I)}$$

respectively, so they have the same normal distribution with zero mean when the model for investigating σ_b^2 holds. Whether $b_{ij(1)}$ is an outlier (and the other of these b's satisfy the conditions of the model) can be investigated by methods for deciding whether a given observation, supposedly in a random sample from a normal population with known mean, is to be considered an outlier. This approach is usable in any nested design for investigating the random variables that have all the subscripts used in the model. The additional variables for the extended model can have all the subscripts that occur for the other random variables of the nested model being extended.

The level of generalization that can be attained by the presence of one or more additional random variables is related to what is investigated and to the number of variance components investigated (decreases as the number of variance components increases). For example, consider the usual model for a balanced three-way cross classification, where $i = 1, \ldots, I$; $j = 1, \ldots, j$; and $k = 1, \ldots, K$. This model is

$$y_{ijk} = \mu + a_i + b_j + c_k + d_{ij} + f_{ik} + g_{jk} + h_{ijk},$$

with μ a parameter and the other quantities random variables. All the random

variables have zero expectation, the a_i have variance σ_a^2, the b_j a variance of $\sigma_b^2, \ldots,$ the h_{ijk} variance σ_h^2. Also, the random variables are mutually independent.

When all of $\sigma_b^2, \ldots, \sigma_g^2$ are investigated, a single random variable e' occurs in the extension model. When σ_a^2, σ_b^2, and σ_d^2 are investigated, the extension model is

$$y_{ijk} = \mu + a_i + b_j + c_k + d_{ij} + f_{ik} + g_{jk} + h_{ijk} + e_k,$$

so that K additional random variables are present. The increase in generality level with K additional random variables is much greater than that for one additional variable, even if K = 2. When only σ_a^2 is investigated, JK additional random variables e_{jk} can be present, with an appreciable increase in generality level as compared with one additional random variable or K additional variables. Similar remarks apply to investigation of σ_b^2, σ_c^2, σ_g^2, to investigation of σ_c^2, etc.

General results for N-way classifications with variance components models are stated in the next section.

2. General Extended Models

To cover the very general class of situations considered, a notation similar to that of Walsh (1968), for fixed effects models, is used. Explicit investigation procedures may not have been developed yet for some special cases of the general class of "usual" models that is considered. However, the purpose of this paper is limited to showing how such models can be extended.

Virtually all of the usual models for an N-way classification in variance components express an observation y[i(1),...,i(N)] in the form

$$y[i(1),\ldots,i(N)] = \mu + \sum_{t=1}^{N} \sum_{j_1,\ldots,j_t}^{*} A[i(j_1),\ldots,i(j_t)], \qquad (4)$$

where the Σ^* denotes summation over the values of $j_1,\ldots,j_t$ such that $1 \le j_1 < \ldots < j_t \le N$, perhaps subject to restrictions on the joint values of $j_1,\ldots,j_t$. Also, $1 \le i(1) \le J_1;\ldots;\ 1 \le i(N) \le J_N$, perhaps subject to restrictions on the joint values of $i(1),\ldots,i(N)$. Here μ is constant, all of the $A[i(j_1),\ldots,i(j_t)]$ are random variables with zero expectation, and, for fixed $j_1,\ldots,j_t$, the $A[i(j_1),\ldots,i(j_t)]$ have the same variance $\sigma(j_1,\ldots,j_t)^2$. The only $\sigma(j_1,\ldots,j_t)^2$ investigated are those for which $t \le N-1$.

Sometimes the allowable combinations of $j_1,\ldots,j_t$ for $t \le N-1$ do not include all of $1,\ldots,N$ (such as for nested classifications). Then, a model for $y[i(1),\ldots,i(N)]$ may not be called on N-way design, although it would be a special case of model (4).

The statistics used to investigate $\sigma(j_1,\ldots,j_t)^2$ for given $j_1,\ldots,j_t$ nearly always are also used to investigate all of the $\sigma[H(j_1,\ldots,j_t)]^2$, where $H(j_1,\ldots,j_t)$ denotes any subset of $j_1,\ldots,j_t$; the variance for the random variables involving this subset of $j_1,\ldots,j_t$ is represented by $\sigma[H(j_1,\ldots,j_t)]^2$. Only models (4) where this is the case are considered.

The statistics virtually always used for investigating $\sigma(j_1,\ldots,j_t)^2$ and the $\sigma[H(j_1,\ldots,j_t)]^2$ are sums of squares for differences of observations. These sums of squares virtually always have the property that A's not involving any of $j_1,\ldots,j_t$ cancel out in all of these sums of squares. In fact, one or more nuisance parameters would be introduced if not all of these A's cancelled out in the sums of squares. Only models (4) where the A's have this property for the statistics used are considered.

Now, consider the extended form of model (4) when model (4) satisfies the conditions of the preceding two paragraphs. The most elementary case is that where the variance components investigated are $\sigma(k_1,\ldots,k_u)^2$ and the $\sigma[H(k_1,\ldots,k_u)]^2$, for given values of $k_1,\ldots,k_u$, $(1 \le u \le N-1)$. In this extended

model, y[i(1),...,i(N)] is expressed as

$$\mu + \sum_{t=1}^{N} \sum_{j_1,\ldots,j_t}^{*} A[i(j_1),\ldots,i(j_t)] + e[i(k_1'),\ldots,i(k_{n-u}')],$$

where $k_1',\ldots,k_{N-u}'$ consists of the values of $1,\ldots,N$ that are not equal to any of $k_1,\ldots,k_u$. The $A[i(j_1),\ldots,i(j_t)]$ have the same properties as for model (4), including the conditions imposed. The joint distribution of the $e[i(k_1'),\ldots,i(k_{N-u}')]$ is arbitrary. Also, any permissible dependence between the $e[i(k_1'),\ldots,i(k_{N-u}')]$ and the $A[i(j_1),\ldots,i(j_t)]$ can occur. The first or second moments do not necessarily exist for the e's, and the covariances of the e's and the A's do not necessarily exist. An e[ ] can have a different dependence with each of the A's. However, when the covariance matrix for the y[i(1),...,i(N)] exists, it is positive definite.

Although the $e[i(k_1'),\ldots,i(k_{N-u}')]$ occur in the extended model, they cancel out in the statistics used for investigating $\sigma(k_1,\ldots,k_u)^2$ and the $\sigma[H(k_1,\ldots,k_u)]^2$. Thus, an investigation based on model (4) remains applicable when this extension of model (4) represents the observations.

The general case is where the $\sigma(k_1^{(r)},\ldots,k_{u(r)}^{(r)})^2$ and the $\sigma[H(k_1^{(r)},\ldots,k_{u(r)}^{(r)})]^2$ are investigated for $r = 1,\ldots,R$. Here, $u(r) \le N - 1$ for all r. In this general extended model, y[i(1),...,i(N)] is expressed as

$$\mu + \sum_{t=1}^{N} \sum_{j_1,\ldots,j_t}^{*} A[i(j),\ldots,i(j_t)] + e[i(s_1),\ldots,i(s_p)],$$

where $s_1,\ldots,s_p$ consists of the values of $1,\ldots,N$ (if any) that are not equal to any of the $k_1^{(r)},\ldots,k_{u(r)}^{(r)}$ for $r = 1,\ldots,R$.

As for the elementary case, the $e[i(s_1),\ldots,i(s_p)]$ cancel out in the statistics used for the investigation. Also, the comments about the e[ ] for the

elementary case again apply. The value of p may be small and can be zero.

CHAPTER V

EXTENSIONS OF MODELS HAVING FIXED EFFECTS, WITH INVESTIGATION OF ALL EFFECTS

ONE-WAY ANOVA MODEL

0. Summary

Consider the standard one-way ANOVA model with fixed effects. Extensions are made of this one-way model by addition of further "error" terms of one or two kinds. For the extensions, exact procedures are obtained for investigating all the effects that appear in the standard model, and for investigating subsets of these effects. For several extensions, the customary results for the standard model remain applicable (and are of the nature of the material in Chapter III). Some procedures differing from the customary ones are used for the other extensions.

1. Introduction

The balanced fixed effects model for one-way analysis of variance is

$$y_{jk} = \mu + \alpha_j + e_{jk}, \tag{1}$$

where $j = 1,\dots,J$ and $k = 1,\dots,K$, with $J,K \geq 2$. Here, y_{jk} is an observed random variable, μ is a parameter, α_j is a parameter such that $\alpha_1 + \dots + \alpha_J = 0$, and e_{jk} is an unobserved random variable. The e_{jk} are assumed to be uncorrelated with zero expectation and the same positive variance σ^2. They are also assumed to have a joint normal distribution when something other than a point estimate is desired.

Seven extensions of the standard model (1) are given. An extension is made by adding more "error" terms, of one or two kinds, to the standard model for y_{jk}. An extension occurs such that μ, σ^2, and one or more of the α can all be investigated by exact procedures. Some of the statistics for these investigations differ from those customarily used for model (1).

2. Extended Models

The extensions given depend on which effects are to be simultaneously investigated and on whether the statistical procedures are restricted to those customarily used for model (1). The results using standard procedures are given here, rather than in Chapter III, for presentation convenience.

First, consider the extended model when μ, σ^2, and one or more of the α_j are all to be investigated. Limitation to the results customarily used does not apply to this case. The model is

$$y_{jk} = \mu + \alpha_j + e_{jk} + e'_k, \qquad (2)$$

where μ, α_j, and e_{jk} have the same properties as for model (1). The additional random errors $e'_1, \ldots, e'_K$ must sum to zero.

Second, consider the case where the eligible procedures are not restricted and investigation of σ^2 and one or more of the α_j is to occur. The extended model for this case is

$$y_{jk} = \mu + \alpha_j + e_{jk} + e^*_k. \qquad (3)$$

Model (3) is an extension of model (2).

Third, consider the case where the procedures are restricted to these customarily used for model (1) and investigation of σ^2 and one or more of the α_j is to occur. The extension for this case is

$$y_{jk} = \mu + \alpha_j + e_{jk} + e'. \qquad (4)$$

This is the least general of the extensions considered and, for $K > 2$, is much less general than model (2).

Fourth, consider the case where the eligible procedures are not limited and both μ and σ^2 are investigated. The extended model is

$$y_{jk} = \mu + \alpha_j + e_{jk} + e'_k + e''_j, \qquad (5)$$

where $e'_1 + \ldots + e'_K = 0$ and $e''_1 + \ldots + e''_J = 0$. Model (5) is an extension of models (2) and (4).

Fifth, consider the case where the procedures are restricted to the customary ones for model (1) and investigation of both μ and σ^2 occurs. The extension is

$$y_{jk} = \mu + \alpha_j + e_{jk} + e_j'' , \tag{6}$$

where $e_1'' + \ldots + e_J'' = 0$. Model (5) is also an extension of this model.

Only σ^2 is investigated for the final two extensions. When the eligible procedures are not restricted, the extension is

$$y_{jk} = \mu + \alpha_j + e_{jk} + e_k^* + e_j^{**} . \tag{7}$$

This is the most general model considered and is an extension of model (5).

Finally, suppose that the procedures are limited to those customarily used and σ^2 is investigated. The extended model is

$$y_{jk} = \mu + \alpha_j + e_{jk} + e_j^{**} . \tag{8}$$

Model (8) has substantially less generality than model (7) but is the most general extension given that is usable with an application of the method for rejection of outliers that is described in earlier material.

As is easily seen, the generality of a model is strongly reduced when the eligible procedures are limited to those customarily used for model (1). Actually, the only statistic encountered that is not standard is the statistic for investigating σ^2 (also used in tests and confidence regions for the other effects).

3. Basis for Investigations

This section states the statistics considered for use along with the effects they investigate, the extended model(s) for which the additional terms do not occur, and pertinent properties. For a given investigation, at least one statistic is introduced for each type of effect (μ, σ^2, one or more of the α_j) that is to be investigated. A statistic for investigating σ^2 is always included, since this statistic occurs in the tests and confidence intervals for any type of effect, and

also occurs in estimates of variances for point estimates of effects.

Some of the more elementary probability properties of the statistics are stated without verification. However, proofs are easily obtained from considerations such as those given in Graybill (1961). Also, the customary results are obtained from material in Graybill (1961).

Some further notation is introduced for stating the statistics that are considered for possible use.

$$\hat{\mu} = y_{..}, \qquad \hat{\alpha}_j = y_{j.} - y_{..},$$

$$s_a^2 = \sum_{j=1}^{J} \hat{\alpha}_j^2/(J-1), \quad s_I^2 = \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{jk} - y_{j.})^2/J(K-1),$$

$$s_{II}^2 = \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{jk} - y_{j.} - y_{.k} + y_{..})^2/(J-1)(K-1),$$

$$F_{aI} = Ks_a^2/s_I^2 \qquad F_{aII} = Ks_a^2/s_{II}^2.$$

The statistic s_I^2 is the customary unbiased estimate of σ^2 for model (1) and is free of the additional random terms for extended models (4), (6), and (8). The statistic s_{II}^2 is an unbiased estimate of σ^2 and is free of the additional random terms for all the extended models. With models (4), (6), (8) and normality, $J(K-1)s_I^2/\sigma^2$ has a χ^2-distribution with $J(K-1)$ degrees of freedom. For all the models and normality, $(J-1)(K-1)s_{II}^2/\sigma^2$ has a χ^2-distribution with $(J-1)(K-1)$ degrees of freedom.

The statistic s_a^2 is free of the additional random terms for models (2), (3), and (4). When the normality assumption also holds for the e_{jk}, and any of models (2), (3), or (4) applies, the statistic F_{aII} has an F-distribution with $J-1$ and $(J-1)(K-1)$ degrees of freedom under the null hypothesis that the α_j are all zero. This is readily verified by showing that the $y_{j.} - y_{..}$ are uncorrelated with the

$y_{jk} - y_{j.} - y_{.k} + y_{..}$ and that, under the null hypothesis, $K(J-1)s_a^2/\sigma^2$ has a χ^2-distribution with J-1 degrees of freedom and $(J-1)(K-1)s_{II}^2/\sigma^2$ has a χ^2-distribution with (J-1)(K-1) degrees of freedom. For normality, the α_j all zero, and model (4), the statistic F_{aI} has an F-distribution with J-1 and J(K-1) degrees of freedom (the customary result when model (1) applies).

The statistic $\hat{\alpha}_j$ is the customary unbiased estimate of α_j for model (1) and is free of the additional random terms for models (2), (3), and (4). For these extended models, $(J-1)s_{II}^2/JK$ is an unbiased estimate of the variance of $\hat{\alpha}_j$ and, when the normality assumption also holds for the e_{jk}, is independent of $\hat{\alpha}_j$ (since the $y_{j.} - y_{..}$ are uncorrelated with the $y_{jk} - y_{j.} - y_{.k} + y_{..}$). The statistic $(J-1)s_{II}^2/JK$ is an unbiased estimate of the variance of $\hat{\alpha}_j$ when model (4) applies and, if the normality assumption also holds, is independent of $\hat{\alpha}_j$ (the customary results when model (1) applies). The distribution of $\hat{\alpha}_j$ is normal with mean μ and variance $(j-1)\sigma^2/JK$ when the normality assumptiom applies and any of models (2), (3), or (4) holds. These properties can be used to construct t-statistics for investigating linear combinations of the α_j.

Finally, $\hat{\mu}$ is the customary unbiased estimate of μ for model (1) and is free of the additional random terms for models (2), (5), and (6). For these extended models, s_{II}^2/JK is an unbiased estimate of the variance of $\hat{\mu}$ and, when the normality assumptiom also holds for the e_{jk}, is independent of $\hat{\mu}$ (since $y_{..}$ is uncorrelated with $y_{jk} - y_{j.} - y_{.k} + y_{..}$). The statistic s_I^2/JK is an unbiased estimate of the variance of $\hat{\mu}$ when model (6) holds and, if the normality assumption also applies, is independent of $\hat{\mu}$ (the customary model (1) results). The distribution of $\hat{\mu}$ is normal with mean μ and variance σ^2/JK when the normality assumption holds and any of models (2), (5), (6) applies. These properties can be used to construct a t-statistic for investigating μ.

TWO-WAY ANOVA WITH INTERACTIONS

0. Summary

Consider the standard two-way cross-classification model for ANOVA with fixed effects and interaction. This standard model is generalized, in many ways, by addition of one, two, or three more "error" terms that are of different kinds. For these extensions, exact procedures are obtained for investigating all the effects and for investigating various subsets of these effects.

1. Introduction

The balanced fixed effects model for two-way analysis of variance, with cross-classification and interaction, is (Graybill (1961))

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}, \tag{1}$$

where $i = 1,\ldots,I$; $j = 1,\ldots,J$; $k=1,\ldots,K$; with $I, J, K \geq 2$.

Here, y_{ijk} is an observed random variable, μ is a parameter, α_i is a parameter such that $\alpha_1 + \ldots + \alpha_I = 0$, and β_j is a parameter such that $\beta_1 + \ldots + \beta_J = 0$. The γ_{ij} are parameters such that $\gamma_{i1} + \ldots + \gamma_{iJ} = 0$ for all i, and $\gamma_{1j} + \ldots + \gamma_{Ij} = 0$ for all j. The e_{ijk} are unobserved random variables that are mutually uncorrelated with zero mean and variance σ^2. When something besides point estimates is desired, the e_{ijk} are also assumed to have a joint normal distributiom.

Sixteen extensions of the standard model (1) are given, and exact investigation procedures are obtained for all of them. Each extension is made by addition of one, two, or three more "error terms (that are of different kinds) to the standard model for y_{ijk}. One extended model is such that all of μ, σ^2, one or more of the α_i, one or more of the β_j, and one or more of the γ_{ij} can be investigated. Other extensions are given for investigating subsets of these types of effects. Investigatiom procedures customarily used for model (1) are applicable for part, but not all, of an investigation.

2. Extended Models

First, consider the extension where μ, σ^2, at least one of the α_i, at least one of the γ_{ij} are all investigated. The extension is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e'_{jk} + e''_{ik}, \qquad (2)$$

where, as in the remainder of the section, μ, α_i, β_j, γ_{ij}, and e_{ijk} have the same properties as for model (1). The additional random variables e'_{jk} and e''_{ik} satisfy

$$\sum_{k=1}^{K} e'_{jk} = C = -\sum_{k=1}^{K} e''_{ik},$$

for all i and j, where the constant C can have any value.

Second, consider the extension where the types of effects investigated are σ^2, α_i, β_j, γ_{ij}. For this case,

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e'_{jk} + e''_{ik}, \qquad (3)$$

with the additional random variables satisfying

$$\sum_{k=1}^{K} e'_{jk} = C_1, \qquad \sum_{k=1}^{K} e''_{ik} = C_2,$$

for all i and j, where the constants C_1 and C_2 can have any values. Model (3) is an extension of model (2).

Third, consider the extension where the effects investigated are μ, σ^2, α_i, γ_{ij}. This extended model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e'_{jk} + e''_{jk}, \qquad (4)$$

where the additional random variables satisfy (for i = 1,..., I).

$$\sum_{k=1}^{K} e''_{ik} = C, \quad \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (e'_{jk} + e''_{ik}) = 0.$$

Model (4) is an extension of model (2).

Fourth, consider the case where the types of effects investigated are μ, σ^2, β_j, γ_{ij}. The extension is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e'_{jk} + e''_{ik}, \qquad (5)$$

where the additional random variables satisfy (for $j = 1, \ldots, J$)

$$\sum_{k=1}^{K} e'_{jk} = C, \qquad \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (e'_{jk} + e''_{ik}) = 0.$$

Model (5) is an extension of model (2).

Fifth, the types of effects investigated are σ^2, α_i, γ_{ij} and the extended model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e'_{jk} + e''_{ik}, \qquad (6)$$

where the additional random variables e''_{ik} satisfy

$$\sum_{k=1}^{K} e''_{ik} = C, \qquad (i=1, \ldots, I).$$

Model (6) is an extension of model (4).

Sixth, σ^2, β_j, γ_{ij} are types of effects investigated and the extension is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e'_{jk} + e''_{ik}, \qquad (7)$$

where the additional random variables e'_{jk} satisfy

$$\sum_{k=1}^{K} e'_{jk} = C, \qquad (j=1, \ldots, J).$$

Model (7) is an extension of model (5).

Seventh, consider the extension where the types of effects investigated are μ, σ^2, γ_{ij}. The model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e'_{jk} + e''_{ik}, \qquad (8)$$

where the additional random variables satisfy

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (e'_{jk} + e''_{ik}) = 0.$$

Model (8) is an extension of models (4) and (5).

Eighth, σ^2 and γ_{ij} are the types of effects investigated and the extension is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e'_{jk} + e''_{ik} . \quad (9)$$

Model (9) is an extension of models (6), (7), and (8).

Ninth, consider the case where the types of effects investigated are μ, σ^2, α_i, β_j. The extended model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e'_{jk} + e''_{ik} + e'''_{ij} , \quad (10)$$

where the additional random variables satisfy (i=1,...,I; j=1,...,J)

$$\sum_{i=1}^{I} e'''_{ij} = C_1, \quad \sum_{j=1}^{J} e'''_{ij} = C_2, \quad \sum_{k=1}^{K} e''_{ik} = C_3,$$

$$\sum_{k=1}^{K} e'_{jk} = C_4, \quad \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (e'_{jk} + e''_{ik} + e'''_{ij}) = 0,$$

where the constants C_1, C_2, C_3, C_4 can have any values, subject to the triple summation. Model (10) is an extension of model (2).

Tenth, σ^2, α_i, β_j are the types of effects investigated and the extended model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e'_{jk} + e''_{ik} + e'''_{ij}, \quad (11)$$

where the additional random variables satisfy (i=1, ..., I; j=1, ..., J)

$$\sum_{i=1}^{I} e'''_{ij} = C_1, \quad \sum_{j=1}^{J} e'''_{ij} = C_2,$$

$$\sum_{k=1}^{K} e''_{ik} = C_3, \quad \sum_{k=1}^{K} e'_{jk} = C_4.$$

Model (11) is an extension of models (3) and (10).

Eleventh, the types of effects investigated are μ, σ^2, α_i and the extension is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e'_{jk} + e''_{ik} + e'''_{ij}, \quad (12)$$

where the additional random variables satisfy (i=1, ..., I)

$$\sum_{j=1}^{J} e'''_{ij} = C_1, \quad \sum_{k=1}^{K} e''_{ik} = C_2, \quad \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (e'_{jk} + e''_{ik} + e'''_{ij}) = 0.$$

Model (12) is an extension of models (4) and (10).

Twelfth, μ, σ^2, β_j are the types of effects investigated and the extended model us

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e'_{jk} + e''_{ik} + e'''_{ij} , \tag{13}$$

where the additional random variables satisfy (j=1, ..., J)

$$\sum_{i=1}^{I} e'''_{ij} = C_1, \quad \sum_{k=1}^{K} e''_{jk} = C_2, \quad \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (e'_{jk} + e''_{ik} + e'''_{ij}) = 0.$$

Model (13) is an extension of models (5) and (10).

Thirteenth, the types of effects investigated are σ^2, α_i, and the extension is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e'_{jk} + e''_{ik} + e'''_{ij}, \tag{14}$$

where the additional random variables satisfy (i=1, ..., I)

$$\sum_{j=1}^{J} e'''_{ij} = C_1, \quad \sum_{k=1}^{K} e''_{ik} = C_2.$$

Model (14) is an extension of models (6) and (12).

Fourteenth, σ^2, β_j are the types of effects investigated and the extended model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e'_{jk} + e''_{ik} + e'''_{ij} , \tag{15}$$

where the additional random variables satisfy (j = 1, ..., J)

$$\sum_{i=1}^{I} e'''_{ij} = C_1, \quad \sum_{k=1}^{K} e'_{jk} = C_2.$$

Model (15) is an extension of models (7) and (13).

Fifteenth, μ and σ^2 are the types of effects investigated and the extended model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e'_{jk} + e''_{ik} + e'''_{ij}, \quad (16)$$

where the additional random variables satisfy

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (e'_{jk} + e''_{ik} + e'''_{ij}) = 0.$$

Model (16) is an extension of models (8), (12), and (13).

Sixteenth, the type of effect investigated is σ^2 and the extension is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} + e'_{jk} + e''_{ik} + e'''_{ij}. \quad (17)$$

Model (17) is an extension of all the other extensions.

3. Basis for Investigation

The statistics considered for use are stated here, along with the parameters they investigate, the extended model(s) for which the additional random terms sum out or cancel out, and their pertinent properties. At least one statistic is introduced for each type of parameter that is investigated. Further statistics may also be introduced for use in estimating variances of estimates and/or for use in tests or confidence regions.

For brevity, some of the more evident probability properties for the statistics are stated without justification. Verification is readily obtained, for example, from material in Graybill (1961). In particular, the customary properties for statistics under model (1) are obtainable from Graybill (1961).

The additional notation for statistics is

$$\hat{\mu} = y_{...}, \quad \hat{\alpha}_i = y_{i..} - y_{...}, \quad \hat{\beta}_j = y_{.j.} - y_{...},$$

$$\hat{\gamma}_{ij} = y_{ij.} - y_{i..} - y_{.j.} + y_{...}, \quad s_a^2 = \sum_{i=1}^{I} (y_{i..} - y_{...})^2,$$

$$s_b^2 = \sum_{j=1}^{J} (y_{.j.} - y_{...})^2, \quad s_g^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij.} - y_{i..} - y_{.j.} + y_{...})^2,$$

$$S_e^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{ijk} - y_{ij.} - y_{i.k} - y_{.jk} + y_{i..} + y_{.j.} + y_{..k} - y_{...})^2,$$

$$F_a = JK(J-1)(K-1)S_a^2/S_e^2, \qquad F_b = IK(I-1)(K-1)S_b^2/S_e^2,$$

$$F_g = K(K-1)S_g^2/S_e^2.$$

The statistic $S_e^2/(I-1)(J-1)(K-1)$ is an unbiased estimate of σ^2 and is free of the additional random variables for all the extended models considered. Customarily, for model (1), the statistic

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{ijk} - y_{ij.})^2. \tag{17}$$

would be used instead of S_e^2. However, the generality level is hugely increased by use of S_e^2, and the degrees of freedom for "error" are only reduced from $IJ(K-1)$ to $(I-1)(J-1)(K-1)$. This is not an important reduction unless I and J are small. When the normality assumption for model (1) also holds, S_e^2/σ^2 has a χ^2-distribution with $(I-1)(J-1)(K-1)$ degrees of freedom for all the extensions.

The statistics S_a^2 and $\hat{\alpha}_i$ are free of the additional random variables for models (2), (3), (4), (6), (10), (11), (12), and (14). For each of these models and normality, JKS_a^2/σ^2 has a χ^2-distribution with I-1 degrees of freedom. Also, under the null hypothesis that all the α_i are zero, F_a has an F-distribution with I-1 and $(I-1)(J-1)(K-1)$ degrees of freedom. This is verified by showing that the $y_{i'..} - y_{...}$ are uncorrelated with all of the $(y_{ijk} - y_{ij.} - y_{i.k} - y_{.jk} + y_{i..} + y_{.j.} + y_{..k} - y_{...})$, which also shows that all the $\hat{\alpha}_i$ are independent of S_e for the normality case. The statistic $\hat{\alpha}_i$ is an unbiased estimate of α_i for the stated models and has variance σ^2/JK, which is estimated by S_e^2/JK. Its distribution is normal for the normality case. These properties, including independence with S_e, provide a basis for constructing t-statistics to investigate specified linear

combinations of the α_i (for the normality case).

The statistics S_b^2 and $\hat{\beta}_j$ are free of the additional random variables for models (2), (3), (5), (7), (10), (11), (13), and (15). With any of these models and normality, IKS_b^2/σ^2 has a χ^2-distribution with J-1 degrees of freedom. When also the null hypothesis that all the β_j are zero holds, F_b has an F-distribution with J-1 and (I-1)(J-1)(K-1) degrees of freedom. This is proved by showing that the $y_{.j'.} - y_{...}$ are uncorrelated with all of the $(y_{ijk} - y_{ij.} - y_{i.k} - y_{.jk} + y_{i..} + y_{.j.} + y_{..k} - y_{...})$, which also proves that all the $\hat{\beta}_j$ are independent of S_e for the normality case. The statistic $\hat{\beta}_j$ is an unbiased estimate of β_j for the stated models and has variance σ^2/IK. When the normality assumption holds, $\hat{\beta}_j$ has a normal distribution. Thus, for normality, t-statistics (using S_e) can be constructed for investigating specified linear combinations of the β_j.

The statistics S_g^2 and $\hat{\gamma}_{ij}$ are free of the additional random variables for models (2) - (9). For these models and normality, KS_g^2/σ^2 has a χ^2-distribution with (I-1)(J-1) degrees of freedom. Also, under the null hypothesis that all the γ_{ij} are zero, F_g has an F-distribution with (I-1)(J-1) and (I-1)(J-1)(K-1) degrees of freedom. This is verified by showing that the $(y_{i'j'.} - y_{i'..} - y_{.j'.} + y_{...})$ are uncorrelated with all the $(y_{ijk} - y_{ij.} - y_{i.k} - y_{.jk} + y_{i..} + y_{.j.} + y_{..k} - y_{...})$, which also proves that all the $\hat{\gamma}_{ij}$ are independent of S_e for the case of normality. The statistic $\hat{\gamma}_{ij}$ is an unbiased estimate of γ_{ij} for models (2) - (9) and has variance σ^2/K. The distribution of $\hat{\gamma}_{ij}$ is normal when the normality assumption is satisfied. Hence, t-statistics, using S_e, can be developed for investigating specified linear combinations of the γ_{ij} for the case of normality.

The statistic $\hat{\mu}$ is free of the additional random variables for each of models (2), (4), (5), (8), (10), (12), (13), and (16). The statistic $\hat{\mu}$ is an unbiased estimate of μ for these models and has variance σ^2/IJK (estimated by S_e^2/IJK). The distribution of $\hat{\mu}$ is normal, and $\hat{\mu}$ is independent of S_e, for the normality case.

This independence follows from zero correlation of $\hat{\mu}$ with all of the $(y_{ijk} - y_{ij.} - y_{i.k} - y_{.jk} + y_{i..} + y_{.j.} + y_{..k} - y_{...})$. Thus, under normality

$$[I(I-1)J(J-1)K(K-1)]^{1/2}(\hat{\mu} - \mu)/S_e$$

has a t-distributiom with (I-1)(J-1)(K-1) degrees of freedom, and can be used for investigating μ.

CHAPTER VI

EXTENSIONS OF MODELS HAVING RANDOM EFFECTS, WITH INVESTIGATION OF ALL EFFECTS

ONE-WAY ANOVA MODEL

0. Summary

Considered is the standard one-way ANOVA with random effects (variance components model). This standard model is generalized in several ways. For these extensions, exact results are developed for investigating the mean and both variance components (of the standard model) and for investigating subsets of these parameters. Customary procedures for the standard model remain usable for some investigations and extensions, but more general models are applicable when use of customary results is not required.

1. Introduction

The balanced random effects (variance components) model for one-way ANOVA is (see Graybill (1961))

$$y_{jk} = \mu + a_j + b_{jk}, \tag{1}$$

where $j = 1,\ldots,J$ and $k = 1,\ldots,K$ with $J,\ K \geq 2$. Here, y_{jk} is an observed random variable, μ is a parameter, the a_j are unobserved random variables with zero mean and the same variance σ_a^2, the b_{jk} are unobserved random variables with zero mean and the same variance σ_b^2, and the a_j, b_{jk} are mutually uncorrelated. When something other than a point estimate is desired, the a_j, b_{jk} are also assumed to have a joint normal distribution.

Five extensions of the standard model (1) are given. Each extension is made by addition of one or two types of "error" terms to the standard model for y_{jk}. One extension is such that all of μ, σ_a^2, and σ_b^2 can be investigated by methods with exactly determined properties. Not all of the statistics for this investigation are the same as those customarily used for model (1). Other extended models

are given for investigating σ_a^2 and σ_b^2, and for investigating σ_b^2. The customary procedures are applicable for some of these extensions but different procedures are used for the others.

2. Extensions

The extended models depend on the parameters to be investigated and on whether the investigation procedures are limited to those customarily used for model (1).

First, consider the extension where all of μ, σ_a^2, and σ_b^2 are investigated. Restriction to the customary procedures does not apply to this case. The extended model is

$$y_{jk} = \mu + a_j + b_{jk} + e_k, \tag{2}$$

where, as in the remainder of the section, μ, a_j, and b_{jk} have the same properties as in model (1). The additional random variables $e_1,\ldots,e_K$ sum to zero.

Second, consider the extension where σ_a^2, σ_b^2 are investigated and the procedures for use are not restricted. For this case,

$$y_{jk} = \mu + a_j + b_{jk} + e'_k. \tag{3}$$

Model (3) is a generalization of model (2).

Third, consider the case where the procedures are limited to those customarily used for model (1) and both σ_a^2 and σ_b^2 are investigated. The extension is

$$y_{jk} = \mu + a_j + b_{ij} + e^*. \tag{4}$$

Model (4) has the lowest generality level of the stated extensions and is much less general than model (2) when $K \geq 3$.

Fourth, consider the extension where the eligible procedures are not limited and only σ_b^2 is to be investigated. The model is

$$y_{jk} = \mu + a_j + b_{jk} + e'_k + e''_j. \tag{5}$$

This is the most general extension considered.

Finally, let the procedures be restricted to those customarily used and

consider investigation of σ_b^2. The extension is

$$y_{jk} = \mu + a_j + b_{jk} + e''_j. \qquad (6)$$

Model (6) is appreciably less general than model (5). However, as can be easily shown, model (6) is the most general of the stated extensions for which the type of method for rejection of outliers that was outlined previously is applicable.

3. Basis for Procedures

This section contains a statement of the statistics considered for possible use, the parameters they investigate, the extended model(s) for which they are usable, and their pertinent properties. For an investigation, at least one statistic occurs for each of the types of parameters investigated.

Some of the more evident probability properties for the statistics are stated with no verification. Justification is easily obtained, for example, from material given in Graybill (1961). Also, the customary properties can be obtained from Graybill (1961).

The additional notation used is

$$\hat{\mu} = y_{..}, \qquad S_a^2 = \sum_{j=1}^{J} (y_{j.} - y_{..})^2,$$

$$S_1^2 = \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{jk} - y_{j.})^2 \qquad S_2^2 = \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{jk} - y_{j.} - y_{.k} + y_{..})^2.$$

$$F_{a1} = JK(K-1) S_a^2/(J-1) S_1^2, \qquad F_{a2}^2 = K(K-1) S_a^2/S_2^2.$$

The statistic $S_1^2/J(K-1)$ is the customary unbiased estimate of σ_b^2 and is free of the additional random variables for models (4) and (6). With these models and the normality assumption (for the a_j and b_{jk}), S_1^2/σ_b^2 has a χ^2-distribution with $J(K-1)$ degrees of freedom. The statistic $S_2^2/(J-1)(K-1)$ is an unbiased estimate of σ_b^2 and is free of the additional random error terms for all of the extended models. For all the models and normality, S_2^2/σ_b^2 has a χ^2-distribution with $(J-1)(K-1)$ degrees of freedom.

The statistic S_a^2 is free of the additional random variables for models (2), (3), and (4). For these models and normality, $KS_a^2/(\sigma_b^2 + K\sigma_a^2)$ has a χ^2-distribution with J-1 degrees of freedom. With model (4) and normality, F_{a1} has an F-distribution with J-1 and J(K-1) degrees of freedom under the null hypothesis of $\sigma_a^2 = 0$, which is the customary result when model (1) applies. For this hypothesis, normality, and any of models (2), (3), or (4), the statistic F_{a2} has an F-distribution with J-1 and (J-1)(K-1) degrees of freedom. This is verified by showing that $y_{j'.} - y_{..}$ is uncorrelated with $y_{jk} - y_{j.} - y_{.k} - y_{..}$ for all j',j,k. The statistic

$$S_a^2(J-1)^{-1} - S_1^2[JK(K-1)]^{-1}$$

is an unbiased estimate of σ_a^2 when model (4) holds, and

$$S_a^2(j-1)^{-1} - S_2^2[K(J-1)(K-1)]^{-1}$$

is an unbiased estimate of σ_a^2 when any of models (2), (3), (4) applies.

Finally, $\hat{\mu}$ is the customary unbiased estimate of μ for model (1) and is free of the additional random terms for model (2). When the normality assumption holds, $\hat{\mu}$ is independent of S_a^2, which is the customary result for model (1). Thus, under normality

$$[J(J-1)]^{1/2}(\hat{\mu} - \mu)/S_a$$

has a t-distribution with J - 1 degrees of freedom, which is a customary result for model (1). Incidentally, the customary results for model (1) are usable when only μ is investigated by use of model (2).

TWOFOLD NESTED ANOVA

0. Summary

Consider the standard twofold nested design for analysis of variance with random effects (a variance components model). For these extensions, exact procedures are obtained for investigating all of the mean effect and the three types of

variance components, and for investigating various subsets of these parameters. Most of the investigation procedures are different from those customarily used for the standard model.

1. Introduction

The balanced twofold nested analysis of variance model with random effects is (Graybill (1961))

$$y_{ijk} = \mu + a_i + b_{ij} + c_{ijk}, \tag{1}$$

where $i = 1,\ldots,I$; $j = 1,\ldots,J$; $k = 1,\ldots,K$; with $I,J,K \geq 2$.

Here, y_{ijk} is an observed random variable, μ is a parameter, the a_i are unobserved random variables with zero mean and variance σ_a^2, the b_{ij} are unobserved random variables with zero mean and variance σ_b^2, the c_{ijk} are unobserved random variables with zero mean and variance σ_c^2, and the random variables are mutually uncorrelated. When significance tests or confidence regions are desired, the random variables are also assumed to have a joint normal distribution.

Seven extensions of model (1) are given, and exact investigation procedures are obtained for all of them. Each extended model is obtained by addition of one, two, or three random terms (of an "error" nature) to the righthand side of model (1). One extension is such that all of μ, σ_a^2, σ_b^2, σ_c^2 can be investigated simultaneously. Other extensions are given for investigating various subsets of these parameters. Investigation procedures customarily used for model (1) are usable in some cases, but many of the procedures are not customarily used for the standard model.

2. Extended Models

First, consider the extension for the case where all of μ, σ_a^2, σ_b^2, σ_c^2 are investigated. The model is

$$y_{ijk} = \mu + a_i + b_{ij} + c_{ijk} + e'_{ik} + e''_{jk}, \tag{2}$$

where, as in all the extensions, μ, the a_i, the b_{ij}, and the c_{ijk} have the same

properties as in model (1). The additional random variables satisfy

$$\sum_{k=1}^{K} e'_{ik} = C = -\sum_{k=1}^{K} e''_{jk},$$

for all i and j, where the constant C can have any value.

Second, consider the extension where σ_a^2, σ_b^2, σ_c^2 are investigated. The extended model is

$$y_{ijk} = \mu + a_i + b_{ij} + c_{ijk} + e'_{ik} + e''_{jk}, \tag{3}$$

where the additional random variables satisfy

$$\sum_{k=1}^{K} e'_{ik} = C_1, \qquad \sum_{k=1}^{K} e''_{jk} = C_2,$$

for all i and j, where the constants C_1 and C_2 can have any values. Model (3) is an extension of model (2).

Third, σ_b^2 and σ_c^2 are the parameters investigated and the extended model is

$$y_{ijk} = \mu + a_i + b_{ij} + c_{ijk} + e'_{ik} + e''_{jk}, \tag{4}$$

where the additional random variables satisfy

$$\sum_{k=1}^{K} e''_{jk} = C, \qquad (j=1,\ldots,J).$$

Model (4) is an extension of model (3).

Fourth, only σ_c^2 is investigated and the extension is

$$y_{ijk} = \mu + a_i + b_{ij} + c_{ijk} + e'_{ik} + e''_{jk} + e'''_{ij}. \tag{5}$$

Model (5) is an extension of model (4).

Fifth, μ and σ_a^2 are to be investigated and the extended model is

$$y_{ijk} = \mu + a_i + b_{ij} + c_{ijk} + e^*_{ijk}, \tag{6}$$

where the e^*_{ijk} satisfy

$$\sum_{k=1}^{K} e^*_{ijk} = 0, \qquad (i=1,\ldots,I;\ j=1,\ldots,J).$$

Model (6) is an extension of model (2).

Sixth, only μ is to be investigated and the extension is

$$y_{ijk} = \mu + a_i + b_{ij} + c_{ijk} + e^*_{ijk}, \tag{7}$$

where the e^*_{ijk} satisfy

$$\sum_{j=1}^{J} \sum_{k=1}^{K} e^*_{ijk} = 0, \qquad (i=1,\ldots,I).$$

Model (7) is an extension of model (6).

Finally, only σ_a^2 is to be investigated and the extended model is

$$y_{ijk} = \mu + a_i + b_{ij} + c_{ijk} + e^*_{ijk}, \tag{8}$$

where the e^*_{ijk} satisfy

$$\sum_{k=1}^{K} e^*_{ijk} = C, \qquad (i=1,\ldots,I;\ j=1,\ldots,J).$$

Model (8) is an extension of model (3) and of model (6).

3. Outline of Procedures

Limitation to consideration of seven sets of the parameters is due to the two-fold nested classification. These are the combinations that are meaningful for investigations. For example, investigation of σ_a^2 and σ_b^2 is not meaningful, because the statistic for investigating σ_c^2 is needed for investigation of σ_b^2. Thus, when investigation of σ_a^2 and σ_b^2 is desired, the case considered is investigation of σ_a^2, σ_b^2, σ_c^2.

The statistics for use in the investigation procedures are stated here, along with identification of the parameter(s) they are used to investigate, the extended models for which they apply, and their pertinent properties. For simplicity, some of the more obvious probability properties for the statistics are given without verification. Justification of these stated properties can be obtained, for example, in Graybill (1961).

The extra notation used is

$$\hat{\mu} = y_{...}, \quad s_a^2 = JK \sum_{i=1}^{I} (y_{i..} - y_{...})^2 ,$$

$$s_b^2 = K \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij.} - y_{i..})^2,$$

$$s_c^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{ijk} - y_{ij.} - y_{i.k} + y_{i..} + y_{.j.} + y_{..k} - y_{...})^2.$$

$$F_a = I(J-1) s_a^2/(I-1) s_b^2, \quad F_b = (I-1)(K-1) s_b^2/I s_c^2.$$

The statistic $s_c^2/(i-1)(J-1)(K-1)$ is an unbiased estimate of σ_c^2 and is free of the additional random variables for all the estensions considered. When the normality assumption for model (1) is also satisfied, s_c^2/σ_c^2 has a χ^2-distribution with $(I-1)(J-1)(k-1)$ degrees of freedom for all the extended models.

The statistic s_b^2 is free of the additional random variables for extended models (2), (3), (4), (6), and (8). For these models, $s_b^2/I(J-1)$ is an unbiased estimate of $\sigma_c^2 + K\sigma_b^2$. When normality also holds, $s_b^2/(\sigma_c^2 + K\sigma_b^2)$ has a χ^2-distribution with $I(j-1)$ degrees of freedom. For models (2), (3), (4) and normality, s_b^2 is independent of s_c^2, so that F_b has an F-distribution with $I(J-1)$ and $(I-1)(J-1)(K-1)$ degrees of freedom under the null hypothesis that $\sigma_b^2 = 0$. Independence of s_b^2 and s_c^2 is verified by showing that the $y_{i'j'.} - y_{i'..}$ are uncorrelated with all of the $(y_{ijk} - y_{ij.} - y_{i.k} - y_{.jk} + y_{i..} + y_{.j.} + y_{..k} - y_{...})$. The statistic

$$[s_b^2/I(J-1) - s_c^2/(I-1)(J-1)(K-1)]/K$$

is an unbiased estimate of σ_b^2 for models (2), (3), and (4).

The statistic s_a^2 is free of the additional random variables for models (2), (3), (6), (7), and (8). For these models, $s_a^2/(I-1)$ is an unbiased estimate of $\sigma_c^2 + K\sigma_b^2 + JK\sigma_a^2$. When normality also holds, $s_a^2/(\sigma_c^2 + K\sigma_b^2 + JK\sigma_a^2)$ has a χ^2-distribution with I-1 degrees of freedom and is independent of s_b^2 (a customary

result). Thus, under the null hypothesis that $\sigma_a^2 = 0$, the statistic F_a has an F-distribution with I - 1 and I(J-1) degrees of freedom. The statistic

$$[s_a^2/(I-1) - s_b^2/I(J-1)]/JK$$

is an unbiased estimate of σ_a^2 for extensions (2), (3), (6), and (8).

Finally, the statistic $\hat{\mu}$ is free of the additional random variables for models (2), (6), and (7). For these models, $\hat{\mu}$ is an unbiased estimate of μ and the variance of $\hat{\mu}$ is

$$(\sigma_c^2 + K\sigma_b^2 + JK\sigma_a^2)/IJK,$$

which is unbiasedly estimated by the statistic $s_a^2/I(I-1)JK$. When the normality assumption also holds, $\hat{\mu}$ is independent of s_a^2 (a customary result) and the statistic

$$[I(I-1)JK]^{1/2}(\mu-\mu)/s_a$$

has a t-distribution with I-1 degrees of freedom.

CHAPTER VII

EXTENSIONS FOR ANALYSIS OF COVARIANCE MODEL AND MIXED MODEL

ONE-WAY ANALYSIS OF COVARIANCE MODEL

0. Summary

Considered are extensions of the standard one-way classification model for analysis of covariance that has fixed effects. For these extensions, nearly all of the customary investigation procedures are shown to have the same properties as for the standard model. An extended model can be developed for a case where, in the statistic(s) used for the investigation, the observations occur exclusively in the form of their differences. An application to rejection of outlying observations is discussed.

2. Introduction

The standard model for the balanced one-way classification in analysis of covariance is

$$y_{jk} = \mu + \alpha_j + \beta x_{jk} + e_{jk}, \tag{1}$$

where $j = 1,\ldots,J$ and $k = 1,\ldots,K$, with $J,K \geq 2$. Here, y_{jk} is an observed variable, x_{jk} is a concomitant variable with fixed known value, μ and β are parameters, α_j is a parameter such that $\alpha_1 + \ldots + \alpha_J = 0$, and e_{jk} is an unobserved random variable. The x_{jk} are such that there is at least one value of j for which $x_{j1},\ldots,x_{jk}$ are not all equal. The e_{jk} are assumed to be uncorrelated with zero expectation and the same positive variance σ^2. They are also assumed to have a joint normal distribution when something other than a point estimate is desired.

A basis for rejection of outlying observations is outlined first. Suppose that, without knowledge of the observed values, y_{jk*} is chosen to be investigated as a possible outlier. On the basis of the known values for $x_{j1},\ldots,x_{jk}$, but without knowledge of the observation values, divide $y_{j1},\ldots,y_{jK}$ into sets of size three and

also zero, one, or two sets of size four (with y_{jk*} not in a set of size four). Any set of size four is converted to a set of size three. This is accomplished by adding two of its observations and dividing this sum by $\sqrt{2}$, to yield one "observation." This "observation" is designated as $y_{jk'}$, where k' is the smaller of the values of k for the two observations added. For each set (now all of size three), develop a statistic of the form

$$y_{jk(1)} + ay_{jk(2)} + by_{jk(3)}, \qquad (2)$$

where

$$1 + a + b = 0$$

$$x_{jk(1)} + ax_{jk(2)} + bx_{jk(3)} = 0,$$

and $k(1) = k^*$ for the set containing y_{jk*}. The division into sets, the conversion of sets of size four (if any), and the choice of k(1), k(2), k(3) within a set, is such that solutions exist for a, b and also the statistic (2) for the set containing y_{jk*} has the smallest variance, subject to $k(1) = k^*$.

Let a = A and b = B for the statistic that includes y_{jk*} and multiply the statistic (2) for any other set by

$$(A^2 + B^2 + 1)(a^2 + b^2 + 1)^{-1}.$$

The resulting statistics do not involve μ, β, or the α_j. They are mutually uncorrelated with zero mean and have the same variance. With normality assumed, the statistic containing y_{jk*} can be investigated by a procedure for examining whether a specified (without knowledge of the observation values) observation, supposedly in a sample from a normal population with zero mean, is an outlier.

The purpose of this paper is to show that nearly all of the exact results developed for model (1) remain exact when this model is generalized. Three kinds of extended models are considered.

2. Extended Models

The first extension of model (1) is for the case where β, σ^2 and one or more of the α_j are investigated. Then, the extended model is

$$y_{jk} = \mu + \alpha_j + \beta x_{jk} + e_{jk} + e', \tag{3}$$

where, as for all extensions, μ, α_j, β, σ^2, x_{jk}, and e_{jk} have the same properties as for model (1).

Next, consider the generalization that occurs when only β and σ^2 are to be investigated. This is also the model to be adopted when the procedure outlined for investigating the existence of an outlier is used. The extended model for these purposes is

$$y_{jk} = \mu + \alpha_j + \beta x_{jk} + e_{jk} + e'_j. \tag{4}$$

Model (4) is an extension of model (3).

Finally, consider the generalization when only one or more of the α_j are to be investigated. This extended model is

$$y_{jk} = \mu + \alpha_j + \beta x_{jk} + e_{jk} + e''_k. \tag{5}$$

Model (5) is also an extension of model (3).

3. Listing of Statistics

Some further notation is first introduced to aid in a listing of statistics that are customarily used to investigate β, σ^2, and one or more of the α_j when model (1) is assumed. These statistics, and their properties under model (1), can be found, for example, in Graybill (1961). Examination of the statistics used for simultaneous investigation of one or more of the types of parameters, as considered in each of the three extensions, shows that no additional error term remains in any statistic.

The notation introduced is

$$E_{xx} = \sum_{j=1}^{J} \sum_{k=1}^{K} (x_{jk} - x_{j.})^2, \qquad T_{xx} = \sum_{j=1}^{J} \sum_{k=1}^{K} (x_{j.} - x_{..})^2,$$

$$E_{yy} = \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{jk} - y_{j.})^2, \quad T_{yy} = \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{j.} - y_{..})^2,$$

$$E_{xy} = \sum_{j=1}^{J} \sum_{k=1}^{K} (x_{jk} - x_{j.})(y_{jk} - y_{j.}),$$

$$T_{xy} = \sum_{j=1}^{J} \sum_{k=1}^{K} (x_{j.} - x_{..})(y_{j.} - y_{..}),$$

which is used in expressing the statistics considered.

All statistics used for investigation of σ^2 are proportional to

$$\hat{\sigma}^2 = (E_{yy} - E_{xy}^2/E_{xx})[K(J-1) - 1]^{-1}.$$

The statistics for investigation of β are

$$\hat{\beta} = E_{xy}/E_{xx}$$

and

$$F = \hat{\beta}^2 E_{xx}/\hat{\sigma}^2.$$

The statistic

$$\frac{1}{(K-1)\hat{\sigma}^2} \left[T_{yy} + E_{yy} + \frac{(T_{xy} + E_{xy})^2}{T_{xx} + E_{xx}} - \left(E_{yy} - \frac{E_{xy}^2}{E_{xx}}\right)\right]$$

is used to test $\alpha_1 = \ldots = \alpha_J = 0$. Investigation of individual α_j, or linear combinations of them, can be developed from use of

$$\hat{\alpha}_j = y_{j.} - y_{..} - \hat{\beta}(x_{j.} - x_{..}).$$

Tests and confidence intervals for linear combinations of the α_j are based on $\hat{\sigma}^2$ and the $\hat{\alpha}_j$ corresponding to the α_j occurring in the linear combination.

SCHEFFÉ TWO-WAY MIXED MODEL FOR ANOVA

0. Summary

Considered is the mixed model for two-way ANOVA that was introduced by Scheffé, including a joint normality assumption when tests or confidence regions are desired.

This model is extended in several ways. For these extensions, the effects ordinarily considered for the Scheffé model can still be investigated. Some of the investigation procedures differ from those developed by Scheffé.

1. Introduction and Discussion

The Scheffé mixed model for ANOVA can be stated in the form

$$y_{ijk} = (\mu + \alpha_i) + b_j + c_{ij} + e_{ijk}, \qquad (1)$$

where $i = 1,\dots,I$; $j = 1,\dots,J$; $k=1,\dots,K$; with $I,J,K \geq 2$.

Here y_{ijk} is an observed random variable, μ is an unknown parameter, and the α_i are unknown parameters such that $\alpha_1 + \dots + \alpha_I = 0$. The b_j are unobserved random variables with zero expectation that are uncorrelated and have the same unknown variance σ_B^2. The c_{ij} are unobserved random variables with zero expectation and such that

$$\sum_{i=1}^{I} c_{ij} = 0, \qquad (j = 1,\dots,J).$$

The unknown value of $\mathrm{var}(c_{ij})$ is the same for all j and is denoted by σ_{ii}, which a parameter of the set $\sigma_{ii'}$, $(i,i' = 1,\dots,I)$, with $||\sigma_{ii'}||$ symmetric and positive indefinite. Also, $\mathrm{cov}(c_{ij},c_{i'j'}) = 0$ for $j \neq j'$ and

$$\mathrm{cov}(c_{ij},c_{i'j}) = \sigma_{ii'} - \sigma_{i.} - \sigma_{.i'} + \sigma_{..},$$

where

$$\sigma_{i.} = \sum_{i'=1}^{I} \sigma_{ii'}/I, \quad \sigma_{.i'} = \sum_{i=1}^{I} \sigma_{ii'}/I, \qquad \sigma_{..} = \sum_{i=1}^{I} \sigma_{i.}/I.$$

In addition,

$$\mathrm{cov}(b_j,c_{ij}) = \sigma_{i.} - \sigma_{..}$$

and $\mathrm{cov}(b_j,c_{ij'})$ is zero for $j \neq j'$. Finally, the e_{ijk} are unobserved random variables of an error nature that have zero expectation and the same unknown variance σ_e^2. The e_{ijk} are mutually uncorrelated and are uncorrelated with the b_j and the

c_{ij}.

When tests or confidence regions are considered, the b_j, the c_{ij}, and the e_{ijk} are assumed to have a joint normal distribution.

The effects investigated are one or more of the $\mu_i (=\mu+\alpha_i)$, one or more of the α_i including

$$\sigma_A^2 = (I-1)^{-1} \sum_{i=1}^{I} \alpha_i^2,$$

σ_B^2, one or more of the $\sigma_{ii'}$,

$$\sigma_{AB}^2 = (I-1)^{-1} \sum_{i=1}^{I} \sigma_{ii'},$$

and σ_e^2. Of course, σ_{AB}^2 can be investigated by procedures for investigating one or more of the $\sigma_{ii'}$. Scheffé gives procedures for investigating these parameters in Scheffé (1956) and (1959).

Ten extensions of model (1) are given. Each extended model is formed by addition of one, two, or three more random error terms to the righthand side of equation (1). One extension is such that one or more of the μ_i, one or more of the α_i, σ_B^2, one or more of the $\sigma_{ii'}$, and σ_e^2 can be simultaneously investigated. Appropriate subsets of these types of parameters can be investigated for the other extensions. The procedures developed by Scheffé for model (1) are applicable in some cases but a different "error sum of squares" s_e^2 is used for other cases. The degrees of freedom associated with s_e^2 are $(I-1)(J-1)(K-1)$ while $IJ(K-1)$ degrees of freedom are associated with the error sum of squares used by Sheffé. Use of s_e^2 is the only way the procedures for the extensions differ from those developed by Scheffé.

2. Extended Models

First, consider the extended model for the case where all of one or more

the μ_i $(=\mu+\alpha_i)$, one or more of the α_i, one or more of the $\sigma_{ii'}$, $\sigma_B{}^2$, and $\sigma_e{}^2$ can be simultaneously investigated. This model is

$$y_{ijk} = (\mu + \alpha_i) + b_j + c_{ij} + e_{ijk} + e'_{ik} + e''_{jk}, \qquad (2)$$

where, as for all extended models, μ, the α_i, the b_j, the c_{ij}, and the e_{ijk} have the same properties as for model (1). The extra random errors e'_{ik} and e''_{jk} satisfy

$$\sum_{k=1}^{K} e'_{ik} = C = - \sum_{k=1}^{K} e''_{jk}$$

for all i and j, where the value of the constant C is arbitrary.

Second, consider the extension where the types of parameters considered are the α_i, the $\sigma_{ii'}$, $\sigma_B{}^2$, and $\sigma_e{}^2$. The extension is

$$y_{ijk} = (\mu + \alpha_i) + b_j + c_{ij} + e_{ijk} + e'_{ik} + e''_{jk}, \qquad (3)$$

where the additional random variables satisfy

$$\sum_{k=1}^{K} e'_{ik} = C_1, \qquad \sum_{k=1}^{K} e''_{jk} = C_2,$$

for all i and j, with the constants C_1 and C_2 having arbitrary values. Model (3) is an extension of model (2).

The third extended model involves simultaneous investigation of the α_i, $\sigma_{AB}{}^2$, and $\sigma_e{}^2$. The model is

$$y_{ijk} = (\mu + \alpha_i) + b_j + c_{ij} + e_{ijk} + e'_{ik} + e''_{jk}, \qquad (4)$$

where the e'_{ik} satisfy

$$\sum_{k=1}^{K} e'_{ik} = C.$$

Model (4) is an extension of model (3).

Fourth, consider the extension for investigation of the μ_i, $\sigma_{AB}{}^2$, and $\sigma_e{}^2$. The model is

$$y_{ijk} = (\mu + \alpha_i) + b_j + c_{ij} + e_{ijk} + e'_{ik} + e''_{jk}, \qquad (5)$$

where the e'_{ik} and e''_{jk} satisfy

$$\sum_{k=1}^{K} (e'_{ik} + e''_{jk}) = 0$$

for all i,j. Model (5) is an extension of model (2).

Fifth, the parameters investigated are σ_{AB}^2, σ_e^2 and the extended model is

$$y_{ijk} = (\mu + \alpha_i) + b_j + c_{ij} + e_{ijk} + e'_{ik} + e''_{jk}. \tag{6}$$

Model (6) is an extension of model (5).

Sixth, the μ_i, σ_B^2, and σ_e^2 are investigated. The extension is

$$y_{ijk} = (\mu + \alpha_i) + b_j + c_{ij} + e_{ijk} + e'_{ik} + e''_{jk} + e'''_{ij}, \tag{7}$$

where

$$\sum_{k=1}^{K} e''_{jk} = c_1, \quad \sum_{j=1}^{J} e'''_{ij} = c_2, \quad \sum_{k=1}^{K} (e'_{ik} + e''_{jk} + e'''_{ij}) = 0$$

for all i,j. Model (7) is an extension of model (2).

Seventh, σ_B^2 and σ_e^2 are investigated, and the extended model is

$$y_{ijk} = (\mu + \alpha_i) + b_j + c_{ij} + e_{ijk} + e'_{ik} + e''_{jk} + e'''_{ij}, \tag{8}$$

where

$$\sum_{k=1}^{K} e''_{jk} = c_1, \quad \sum_{j=1}^{J} e'''_{ij} = c_2$$

for all i,j. Model (8) is an extension of model (7).

Eighth, the μ_i and σ_e^2 are investigated and the extension is

$$y_{ijk} = (\mu + \alpha_i) + b_j + c_{ij} + e_{ijk} + e'_{ik} + e''_{jk} + e'''_{ij}, \tag{9}$$

where

$$\sum_{k=1}^{K} (e'_{ik} + \text{[illegible]}' + e'''_{ij}) = 0$$

for all i, j. Model (9) is an extension of model (7).

Ninth, only σ_e^2 is investigated and the extended model is

$$y_{ijk} = (\mu + \alpha_i) + b_j + c_{ij} + e_{ijk} + e'_{ik} + e''_{jk} + e'''_{ij}. \quad (10)$$

Model (10) is an extension of all of models (2) - (9).

Tenth, and last, one or more of the μ_i are investigated and the extension is

$$y_{ijk} = (\mu + \alpha_i) + b_j + c_{ij} + e_{ijk} + e^*_{ijk}, \quad (11)$$

where

$$\sum_{k=1}^{K} e^*_{ijk} = 0$$

for all i, j. Model (11) is an extension of model (9).

3. Basis for Investigation Procedures

Limitation to ten extensions is largely due to the forms that occur for extended models. For example, an extension for investigating the α_i, σ_B^2, and σ_e^2 is of the same form as that for also investigating the $\sigma_{ii'}$. Thus, no model is stated for investigation of the α_i, σ_B^2, and σ_e^2. Also, consideration is limited to combinations of parameter types that are meaningful for investigation using the Scheffé mixed model. For example, consideration of σ_B^2 and/or σ_{AB}^2 is not appropriate, since the statistic for investigating σ_e^2 is needed. Thus, the parameter σ_e^2 is included in any set containing σ_B^2 and/or σ_{AB}^2.

Except in one respect, the investigation procedures considered are those developed by Scheffé (1956) and (1959). The one difference is that the statistic

$$SS_e = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{ijk} - y_{ij.})^2$$

used by Scheffé is replaced by the statistic

$$s_e^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{ijk} - y_{ij.} - y_{i.k} - y_{.jk} + y_{i..} + y_{.j.} + y_{..k} - y_{...})^2$$

at all places where SS_e occurs. The only effect of this replacement is that a sum of squares with (I-1)(J-1)(K-1) degrees of freedom is substituted for a sum of squares with IJ(K-1) degrees of freedom. This change in degrees of freedom is unimportant unless I and J are small. The generality level of the extensions is greatly increased by using s_e^2 instead of SS_e.

For the case of joint normality, s_e^2 has the same independence properties with the other statistics as does SS_e. This is verified examining the correlation properties of the $(y_{ijk} - y_{ij.} - y_{i.k} - y_{.jk} + y_{i..} + y_{.j.} + y_{..k} - y_{...})$ with the $y_{i!.} - y_{...}$, the $y_{.j!} - y_{...}$, the $(y_{ij!} - y_{i!.} - y_{.j!} + y_{...})$, etc.

For a given extended model, verification is obtained by examining the statistics developed by Scheffé (with SS_e replaced by s_e^2) for investigating the parameters considered. It is easily shown that in every case the additional random error terms cancel out or sum out in the statistics of Scheffé (1956) and (1959).

REFERENCES

Graybill, Franklin A. (1961). An Introduction to Linear Statistical Models, Vol. 1. McGraw-Hill. New York.

Guttman, Irwin and Dennis E. Smith. (1969). Investigation of rules for dealing with outliers in small samples from normal distribution: I: Estimation of the mean. Technometrics, 11: 527-550.

Scheffé, Henry. (1956). A "mixed" model for the analysis of variance. Ann. Math. Stat., 27: 23-36.

Scheffé, Henry. (1959). The Analysis of Variance. John Wiley and Sons, New York. 261-274.

Walsh, John E. (1947). Concerning the effect of intraclass correlation on certain significance tests. Ann. Math. Stat., 18: 88-96.

Walsh, John E. (1968). Handbook of Nonparametric Statistics III: Analysis of Variance. D. Van Nostrand Co., Princeton, N. J.